VOLUME XXIX JUNE 1937 PARTS I AND II

BIOMETRIKA

A CONTRIBUTION TO ESKIMO CRANIOLOGY BASED ON PREVIOUSLY PUBLISHED MEASUREMENTS

BY G. M. MORANT, D.Sc.

1. *Introduction.* Among modern races of man the physical type of the Eskimos is undoubtedly one of the most specialized known. The available metrical data relating to it are far more extensive for the cranium than for any other part of the skeleton, or for the living people. Numerous studies of series of Eskimo skulls have been published in the last sixty years, but the majority of these are of little value for statistical purposes, either on account of the fact that the measurements given are inadequate, or because the series described are too small, or for both these reasons. Until recent years the monumental *Crania Groenlandica* of Professors Fürst and Hansen, published in 1915, was by far the most valuable of the contributions to the subject, but it relates only to Eastern Eskimos. The publication by Dr Hrdlička, between 1924 and 1929, of data relating to a large number of Western and Central Eskimo skulls satisfied a long-felt need. These two sets of material are the only ones dealt with, apart from comparative material for other races, in the present paper. The measurements discussed were only treated by the crudest statistical methods when first presented.

2. *Measurements of Eskimo Skulls provided by Dr Aleš Hrdlička.* Dr Hrdlička has published several contributions to Eskimo craniology, but when considering a statistical treatment of his material it is only necessary to refer to two of these:

(*a*) "Catalogue of Human Crania in the United States National Museum Collections", *Proceedings of the United States National Museum*, LXIII (1924), pp. 1–51. This, the first part of the extensive catalogue, contains individual measurements of a number of series of Eskimo skulls from different localities. There are only two of the male series sufficiently long for statistical purposes, viz. one of 40 Greenland Eskimo crania and another of 159 Eskimo crania from the north coast of St Lawrence Island in the Bering Sea. One of the former came from the Noursoak Peninsula on the west coast of Greenland, and there are no recorded localities for the other specimens in the series. No particulars regarding the discovery or age of the material are provided.

(*b*) "Anthropological Survey in Alaska", *Forty-Sixth Annual Report of the Bureau of American Ethnology*, 1928–9, pp. 19–374. This provides (pp. 254–99) a detailed discussion and mean measurements of a considerable number of groups of Eskimo crania, the majority being made up by small numbers of specimens. All the material of this kind previously described by the author is apparently

included here, and there is a considerable amount of new material. The specimens appear to be of modern or recent date, but one series discussed at greater length than the others (pp. 318–29) is believed to represent a population which lived near Point Barrow, on the north coast of Alaska, before contact with Europeans was established. This is known as the "Old Igloos" series. Dr Hrdlička says that for the purpose of this report he re-measured all the specimens which he had previously described. The means for the Greenland and St Lawrence Island series differ somewhat from those in the 1924 *Catalogue*, and the numbers on which these means are based were also changed. It is said that the individual measurements for all the material will be given in a part of the *Catalogue* which has not yet been published.

Before making statistical comparison between the different series, it is necessary to comment on the definitions followed in determining the measurements. Dr Hrdlička has published an account of his technique,* which is based on that of the Monaco Congress of 1906 with several modifications. A list of the measurements given for the Eskimo series follows, the numbers preceded by I.A. being those of the International Agreement and the letters those of the biometric technique:

(i) Maximum "glabello-occipital" length: I.A. 1. This is not precisely the same as *L*, defined to be the maximum calvarial length from the glabella in the median sagittal plane, but the two definitions will give readings which are either identical or very close to one another.

(ii) Maximum breadth above the mastoids and roots of zygomae: I.A. 3. This, again, is not precisely the same as *B*, defined to be the maximum transverse diameter on the parietal bones, but the two definitions will almost invariably give identical or closely similar readings.

(iii) Basion-bregma height: I.A. 4, *a*. Although the basion is insufficiently defined, this measurement may be supposed the same as *H'*.

(iv) Cranial capacity: I.A. 24, *c*. Hrdlička determined this with seed by using a method which he describes in detail. It is commonly found that different methods often give appreciably different results.

(v) Upper facial height from nasion to alveolar point: I.A. 12, *G'H*. The alveolar point is defined to be the "lowest point of the alveolar border between the two median upper incisors".

(vi) Facial length from basion to alveolar point. Hrdlička gives this definition without comment, and it may be presumed that he used the same alveolar point as in finding the upper facial height, so the measurement is *GL*, assuming that the basions used are the same. He diverges here from the Monaco definition

* "Anthropometry. D. Skeletal Parts: the Skull", *American Journal of Physical Anthropology*, II (1919), pp. 401–28. This article was reprinted in the author's *Anthropometry*, Philadelphia (1920). An English translation of the Monaco report is given in the same volume of the *Journal* and in the book.

(I.A. 10), which specifies that the "alveolar point" used in this case is the "median point of the anterior border of the alveolar arch", i.e. Martin's prosthion.

(vii) Chord from nasion to basion: I.A. 9. This may be supposed the same as *LB*, on the supposition that the same basions are intended.

(viii) Maximum bizygomatic breadth: I.A. 8, *J*.

(ix) Nasal height. All the definitions agree in using the nasion as the superior terminal of this measurement. According to the Monaco technique (I.A. 13) the inferior terminal is "the middle of a line connecting the lowest points of the two nasal fossae", which is inexact as the middle point will normally lie *in* the nasal spine and not on any surface. Hrdlička's practice is to "measure separately to each subnasal point and record the mean", the subnasal points being defined as "the lowest point, on each side, on the lower border of the nasal aperture, i.e. the lowest points anteriorly of the two nasal fossae". This measurement is likely to give such closely similar readings that it may be supposed the same as *NH*, *L*.

(x) Maximum breadth of nasal aperture: I.A. 14, *NB*.

(xi) Orbital breadth: I.A. 16. The terminal of this measurement nearest to the median sagittal plane is the dacryon, or "if the dacryon is obliterated, or in an abnormal situation, take the point where the posterior lacrymal crest meets the inferior border of the frontal". The lateral terminal is "the external border of the orbit, at the point where the transverse axis of the orbit meets the border, and parallel as far as possible to the superior and inferior borders". This is an inadequate definition, since there is no exact way of deciding when the dacryon is in an abnormal situation, and the point sometimes substituted for it normally gives a lesser breadth than that from the true dacryon. Hrdlička follows the Monaco instructions, and only gives data for the mean of the two orbital breadths so obtained. His measurement may be denoted by O_1' (or more precisely by $\frac{1}{2}(O_1'R + O_1'L)$), though it is not exactly the same as the true dacryal breadth.

(xii) Orbital height: I.A. 17. This is the maximum height perpendicular to the breadth and it may be supposed the same as O_2. Hrdlička gives the mean of the heights of the two orbits.

These are the only absolute measurements provided by Dr Hrdlička for Eskimo skulls which are dealt with below. He gives data for five additional measurements determined according to the Monaco definitions, viz. the length and breadth of the "upper alveolar arch" and the chord from basion to subnasal point, for which there is little comparative material; the menton-nasion height, an unreliable measurement owing to the fact that it is influenced by wear of the teeth, and the height of the mandible at the symphysis. The omission of a number of customary measurements—such as the principal arcs, minimum frontal breadth and palatal and foraminal measurements—is to be regretted. The indices and angles in curled brackets in the tables below were obtained from the mean values of the component lengths (indices) or sides of the triangle (angles) instead of from

values for individual skulls. The angles ($N\angle$, $A\angle$ and $B\angle$) are the three of the fundamental triangle of which the sides are $G'H$, GL and LB. Hrdlička gives means for the first, which he calls the facial angle, and some of our means for it differ from his quite markedly.

There is reason to suspect that he modified the ways in which some of his measurements were taken between the times when the data were obtained for the 1924 *Catalogue* and 1928 *Report*. This is suggested by a comparison of the means for the two male series given in the former year with those for the same series given after re-measurement, a number of additional specimens having been added in one case. There is a close agreement between corresponding means except in the case of the following characters:

		C	$G'H$	J	NH	$100\,NB/NH$
Greenland	1924	1560 (34)	74·4 (36)	140·0 (30)	53·4 (39)	42·9 (36)
	1928	1518 (42)	76·1 (46)	140·5 (47)	52·4 (48)	44·3 (48)
St Lawrence Island	1924	1506 (129)	76·6 (144)	140·8 (151)	55·4 (150)	44·6 (150)
	1928	1462 (142)	78·2 (139)	142·0 (148)	54·2 (148)	45·2 (148)

The differences may be partly due to the fact that the corresponding means are given for different numbers of specimens, and the sexes of some of them may have been changed, but it is impossible to avoid the conclusion that the divergences for these characters are due primarily to a change in the definitions of the measurements. It appears that no other hypothesis can explain why the facial heights increased on re-measurement while the nasal heights decreased, or why the capacities decreased while the major calvarial chords remained practically unchanged. The differences which must be attributed to personal equation are large enough to be disturbing when an attempt is made to distinguish small differences between neighbouring Eskimo types. The only means of Dr Hrdlička's series used below are those derived from the data given in his 1928 *Report*, but this contains no individual measurements and the standard deviations of two of the series in our Table I were obtained from the readings in the 1924 *Catalogue*.

3. *Measurements of Greenland Eskimo Skulls provided by Professors Carl M. Fürst and F. C. C. Hansen.* The *Crania Groenlandica** of these authors is one of the most valuable and comprehensive treatises of its kind available for any race. Descriptions and detailed individual measurements of 380 crania are given, 14 of these being immature and 8 others unsexed. The sample forms a selection from the total Eskimo population of Greenland, the more densely populated west coast being represented by larger numbers of specimens than the south-west and

* *Crania Groenlandica, A Description of Greenland Eskimo Crania with an Introduction on the Geography and History of Greenland*, Copenhagen (1915).

east coasts. The distributions and means are, unfortunately, for the combined male and female series in the case of all characters except the cephalic index. A biometric treatment of the male series has been given by the present writer.* Only those constants for characters which are available for Hrdlička's series are considered below. Fürst and Hansen's orbital breadth, "from the lateral to the vertical medial margin, which is the direct continuation of the lower orbital margin and which in the Greenlanders is continued sharply and often high on the maxillary bone", may be supposed the same as the biometric orbital breadth O_1, and this is a different measurement from the dacryal breadth given by Hrdlička. There is no doubt as to whether the measurements of the two sets of data may be considered the same in definition or not except in the case of the nasal height. Fürst and Hansen do not define the way in which they determined this measurement, but it may be presumed that it was the same as, or very similar to, that employed by Hrdlička. Their cranial capacities were found with the aid of millet seed and a graduated glass cylinder, and they may also be supposed comparable to his determined by a slightly different technique.

4. *The Variabilities of Male Eskimo Series.* Standard deviations are given in Table I for the only two Eskimo series measured by Hrdlička for which individual measurements have been published (in the 1924 *Catalogue*), for Fürst and Hansen's Greenland series (the constants being taken from the paper cited), and for the long Egyptian series often used for comparative purposes.† These constants have been given for more than twice as many characters relating to the last two series. The data relate to 13 characters and the four series give 6 comparisons in pairs for each, so there is a total of 78 comparisons. In 29 cases the differences exceed three times their probable errors, and several of them are markedly significant, the highest ratio of a difference to its probable error being 9·6. Twelve of the 13 σ's for the St Lawrence Island series are less than the corresponding values for the Egyptian, and the differences exceed three times their probable errors in 8 of these cases: 11 of the σ's for Hrdlička's Greenland series are less than the Egyptian values, and the differences exceed three times their probable errors in the case of 4 of these 11 comparisons. But the σ's for Fürst and Hansen's Greenland series are in excess of the Egyptian in the case of 10 of the 13 characters, and for 4 of these 10 the differences may be supposed significant. These divergences in variability are more marked than those usually found in the comparison of cranial series believed to be racially homogeneous. This is evidently due to the fact that Hrdlička's two series show peculiarly small variation (only one character showing a significant difference in the comparison between them), while the other two show a greater and more common order of variation. It has been found

* In "Studies of Palaeolithic Man. I. The Chancelade Skull and its Relation to the modern Eskimo Skull", *Annals of Eugenics*, I (1926), pp. 257–76.

† Karl Pearson and Adelaide G. Davin, "On the Biometric Constants of the Human Skull", *Biometrika*, XVI (1924), pp. 328–63.

for other material that the populations of small islands—such as the Guanche of Tenerife—are distinctly less variable than others. The distinction of Hrdlička's Greenland series in the respect considered may be due to the fact that the skulls,

TABLE I

*Standard Deviations of Male Series of Crania**

	Eskimo			Egyptian *E*: 26th–30th dynasties (Pearson and Davin)
	St Lawrence Island (Hrdlička)	Greenland (Hrdlička)	Greenland (Fürst and Hansen)	
C	103·6 ±4·4 (129)	86·0 ±7·0 (34)	128·8 ±4·6 (175)	113·5 ±2·0 (753)
L	4·96±·19 (158)	4·36±·34 (38)	5·81±·20 (191)	5·72±·09 (895)
B	4·05±·15 (156)	4·34±·34 (36)	4·52±·16 (191)	4·76±·08 (896)
H'†	4·25±·17 (143)	3·97±·30 (39)	4·79±·17 (183)	5·03±·08 (884)
G'H	3·32±·13 (144)	3·96±·31 (36)	4·39±·15 (191)	4·15±·07 (845)
J	4·86±·19 (151)	5·74±·50 (30)	6·48±·23 (185)	4·57±·08 (785)
NH	2·39±·09 (150)	2·69±·21 (39)	3·10±·11 (192)	2·92±·05 (898)
NB	1·74±·07 (153)	1·67±·13 (36)	1·75±·06 (191)	1·77±·03 (893)
O_2‡	1·62±·06 (148)	1·84±·14 (38)	2·03±·07 (188)	1·88±·03 (888)
O_1'§	1·52±·06 (148)	0·92±·07 (38)	2·46±·09 (189)	1·65±·03 (880)
100 *B*/*L*	2·39±·09 (156)	3·07±·25 (35)	3·00±·10 (190)	2·68±·06 (884)
100 *NB*/*NH*	3·32±·13 (150)	3·74±·30 (36)	3·84±·13 (191)	3·82±·06 (881)
100 O_2/O_1'‖	3·70±·15 (148)	4·83±·37 (38)	5·60±·19 (189)	4·95±·08 (876)

* The ± signs in this paper indicate probable errors, as in all earlier anthropometric papers in *Biometrika*.

† The Egyptian σ is for the vertical height from the basion (*H*) instead of *H'*. Both these measurements are available for the male Eskimo series measured by Fürst and Hansen, the σ's being 4·78±·17 for *H* and 4·79±·17 for *H'*.

‡ The σ's are for the means of the right and left orbital heights in the case of Hrdlička's series, and for the height of the left orbit in the case of the other two series.

§ The σ's are for the means of the right and left orbital breadths from the dacrya (O_1') in the case of Hrdlička's series, and for the maximum breadths from the medial margin and for the left orbit only (O_1 *L*) in the case of the other two series. Both orbital breadths have been given for a few long series and the σ's for them have been found to be in close agreement.

‖ The σ's are for the orbital indices found from the means of the heights and dacryal breadths for the right and left sides (100 O_2/O_1') in the case of Hrdlička's series, and for 100 O_2/O_1, *L* in the case of the other two. These also show a close agreement when found for the same series.

of unknown origin, came entirely or in large part from a small Eskimo community. Fürst and Hansen's series may be considered a sample drawn, more or less at random, from the total Eskimo population of the country. In calculating all the coefficients of racial likeness given in this paper the Egyptian *E* standard deviations were used. These values are probably close to those which would be found for the majority of the series measured by Hrdlička.

5. *Comparisons of Eskimo Series by the Method of the Coefficient of Racial Likeness.* Dr Hrdlička's 1928 *Report* contains individual or mean measurements of 36 groups of male Eskimo skulls ranging in size from one to 153 individuals.

Most of the series are made up by fewer than 10 specimens, and pooling of some of the material is obviously required as a preliminary step towards statistical analysis. Three Western groups were first compiled by combining in two cases the skulls from neighbouring localities in order to make up samples of a sufficient size. The approximate positions of the sites can be seen from the map in Fig. 1, and the numbers in brackets below give the numbers of male skulls from each site. The Western groups are:

W_1—Prince William Sound (1), Kodiak Island (1), Unalaska Peninsula (1), Nushagak Bay (1), Togiak (4), Mumtrak (4), Nelson Island (9), Hooper Bay (9), Lower Yukon and delta (3), Pilot Station, lower Yukon (3), Kotlik and Pastolik (11) and St Michael Island (8).

W_2—St Lawrence Island (153).

W_3—Little Diomede Island (5), and two sites on the mainland of Asia, Indian Point (14) and Puotin (2).

The means for these three groups are given in Table II and they show a remarkably close resemblance for all characters. The crude coefficients of racial likeness are: W_1 and W_2 $-\cdot06 \pm \cdot22$ (18),* W_1 and W_3 $-\cdot49 \pm \cdot23$ (17)† and W_2 and W_3 $-\cdot50 \pm \cdot23$ (17)†. As far as can be seen from the data available, the Eskimo population of the south-west of Alaska, St Lawrence Island and the Asiatic mainland is perfectly homogeneous. Within this area there is no evidence of local populations differing significantly from the prevailing type, though it is quite possible that there are local variants. There is only one skull from Kodiak Island, for example, and it is quite possible that if 50 were available their measurements would distinguish the population from that of St Lawrence Island. At the moment the pooling of the three groups W_1, W_2 and W_3 appears to be justified and the combined means are those of the Western series in Table IV. It will be shown that they are not closely similar to those for any other series available.

Three groups of Eskimo skulls from the north-west and north of Alaska were made up in the following way, these being distinguished from the Western groups because their mean measurements clearly differentiate them:

NW_1—Golovnin Bay (3), Cape Nome (1), Sledge Island (5), Port Clarence (4), Wales (19), Shishmaref (13), Kotzebue (2).

NW_2—Barrow and vicinity (37).

NW_3—Point Barrow (49).

The means for these groups, given in Table II, again show a remarkably close resemblance for all characters. The crude coefficients of racial likeness between them are: NW_1 and NW_2 $-\cdot04 \pm \cdot23$ (17), NW_1 and NW_3 $\cdot56 \pm \cdot23$ (17) and NW_2 and NW_3 $-\cdot18 \pm \cdot23$ (17).† The pooling of the three groups again appears to be fully justified, and accordingly the combined means were computed and they

* For all the characters in Table II except *GL* and $B\angle$.

† For all the characters in Table II except *C*, *GL* and $B\angle$.

are those for the North-Western series in Table IV. It will be shown that they are differentiated from those for all the other series available. It is surprising to find that all the material hitherto considered in this section can be partitioned with so little hesitation into two homogeneous series which are distinctly different from one another. A crude comparison of the means in Table II shows that this

TABLE II

*Mean Measurements of Groups of Male Eskimo Skulls from Alaska**

	Western groups			North-Western groups		
	St Lawrence Island (W_2)†	South-Western (W_1)	Asiatic and Little Diomede Island (W_3)	Kotzebue Sound and north of Norton Sound (NW_1)	Barrow and vicinity (NW_2)	Point Barrow (NW_3)
C	1462 (142)	1503·1 (53)	1470 (5)	1448·4 (40)	—	1324 (5)
L	184·0 (153)	183·3 (55)	185·1 (21)	187·3 (47)	189·0 (37)	187·4 (49)
B	141·9 (153)	142·1 (55)	143·2 (21)	136·6 (47)	137·3 (37)	138·4 (49)
H'	136·8 (145)	135·9 (55)	137·2 (20)	138·0 (45)	137·8 (35)	137·8 (47)
LB	103·6 (145)	103·8 (54)	104·6 (19)	106·5 (45)	106·1 (35)	105·4 (47)
GL	104·3 (131)	103·7 (43)	104·4 (14)	106·0 (39)	103·9 (21)	103·9 (36)
$G'H$	78·2 (139)	78·6 (47)	78·3 (17)	77·6 (39)	78·9 (21)	78·6 (37)
J	142·0 (148)	142·1 (52)	141·9 (21)	141·8 (42)	143·4 (26)	142·6 (44)
NH	54·2 (148)	54·4 (54)	55·0 (21)	54·0 (44)	55·2 (29)	54·8 (46)
NB	24·5 (148)	24·2 (54)	25·0 (21)	23·8 (44)	23·9 (29)	23·1 (46)
O_2	36·8 (145)	36·7 (54)	37·0 (21)	36·3 (44)	36·0 (28)	36·1 (43)
O_1'	40·3 (145)	40·0 (54)	40·6 (21)	40·6 (44)	40·4 (28)	40·2 (43)
$100\ B/L$	77·1 (153)	77·6 (55)	77·4 (21)	73·0 (47)	72·6 (37)	73·9 (49)
$100\ H'/L$	{74·3 (145)}	{74·1 (55)}	{74·1 (20)}	{73·7 (45)}	{72·9 (35)}	{73·5 (47)}
$100\ B/H'$	{103·7 (145)}	{104·6 (55)}	{104·4 (20)}	{99·0 (45)}	{99·6 (35)}	{100·4 (47)}
$100\ NB/NH$	45·2 (148)	44·5 (54)	45·4 (21)	44·2 (44)	43·4 (29)	42·2 (46)
$100\ O_2/O_1'$	91·2 (145)	91·8 (54)	91·1 (21)	89·5 (44)	89·2 (28)	89·9 (43)
$N\angle$	{68°·2 (131)}	{67°·6 (43)}	{67°·8 (14)}	{68°·0 (39)}	{66°·3 (21)}	{66°·7 (36)}
$A\angle$	{67°·0 (131)}	{67°·9 (43)}	{68°·2 (14)}	{69°·2 (39)}	{69°·7 (21)}	{69°·2 (36)}
$B\angle$	{44°·3 (131)}	{44°·5 (43)}	{44°·0 (14)}	{42°·8 (39)}	{44°·0 (21)}	{44°·1 (36)}

* The indices and angles in curled brackets were derived from the means of the chords involved, instead of from values for individual skulls.

† The locations of the groups are shown on the map in Fig. 1.

step is entirely reasonable. In the case of L, B, $100\ B/L$ and $100\ B/H'$ the means for the three Western sub-groups cover a small range and those for the three North-Western sub-groups cover another small range, while there is a clear separation of the two ranges. For C, H', LB, NB, O_2, $100\ H'/L$, $100\ NB/NH$, $100\ O_2/O_1'$ and $A\angle$ the two ranges are also discrete, but the separation between them is less clear. For the remaining characters—GL, $G'H$, J, NH, O_1', $N\angle$ and $B\angle$—the two ranges overlap, but the differences are so small that all between pairs of the six groups are probably insignificant. The two major groups are thus clearly defined and clearly distinguished. It would have been expected from geographical considerations

(see Fig. 1) that the sub-group NW_1 would bear a closer resemblance to the three Western sub-groups than NW_2 and NW_3 would to these three, but there is no suggestion of this from the measurements. There appears in fact to be a distinct cleavage between the types of the south-west and north-west coast of Alaska, though more abundant material would be needed in order to ascertain with precision where the dividing line comes. Some of the series from neighbouring sites making up the sub-groups W_1 and NW_1 are so small that a few might be transferred from one to the other without affecting the position appreciably.

The means of three other series of male Eskimo skulls from Alaska are given by Dr Hrdlička in his 1928 *Report*. They fall within the same area as the Western and North-Western series dealt with above, but they were kept apart from these as their mean measurements evidently differentiate them. The first is made up by 46 specimens from Nunivak Island, the second of 131 from Point Hope and the third is the series of 27 "Old Igloos" skulls from the vicinity of Point Barrow. This last is believed to represent an earlier population than all the others, and it was kept separate on this account, and also because the type it represents is clearly distinguished from all the others determined by Alaskan series. A number of small groups from the islands west of Greenland and the Canadian mainland were pooled to form what will be called the Central Eskimo series. They are: Northern Arctic (5), Melville Peninsula (1), Southampton Island (9), Hudson Bay and Ungava Bay (5), Baffin Land, northern Devon, and vicinity (16) and Smith Sound (7).

The only remaining series measured by Dr Hrdlička is the one of 49 skulls from unknown localities in Greenland. This may be compared with Fürst and Hansen's Greenland Eskimo series. These writers conclude that: "The anthropological characters cannot contribute to a solution of the question of the migration of the Eskimos [in Greenland], owing to the fact that the homogeneity of their anthropological characters clearly shows that the Eskimos of both the west and the east coasts are of the same racial type." This conclusion is based principally on a comparison of the distributions of the measurements for *unsexed* series of crania from different regions of Greenland. Male means computed for three groups into which the total material is divided are given in Table III below in the case of nine of the more important characters. In asking whether the differences between these are significant or not, the standard deviations of Hrdlička's Greenland Eskimo series (given in Table I) were used, and it has been noticed that these are appreciably less than the values for the total series measured by Fürst and Hansen. The only differences greater than three times their probable errors are: C, Eastern less than Western ($\Delta/(\text{p.e. }\Delta) = 5{\cdot}9$) and South-Western (6·3); H', Western less than South-Western (3·3); J, Eastern less than South-Western (3·8); NH, Western less than South-Western (4·0); $100\,NB/NH$, South Western less than Western (3·3). No importance can be attached to any of these differences except those for the capacity. The types appear to differ principally

in size, and it may be noted that the South-Western series has the largest means in the case of all the absolute measurements except *L*.

TABLE III

Mean Measurements for Male Series of Greenland Eskimo Skulls

	Western	South-Western	Eastern	Pooled	Unknown localities
	Fürst and Hansen				Hrdlička
C	1536·2 (95)	1549·0 (48)	1465·8 (32)	1526·8 ±6·6 (175)*	1518 ±9·0 (42)†
L	188·6 (100)	188·5 (54)	187·9 (37)	188·4 ±·28 (191)	189·7±·42 (49)
B	134·5 (100)	135·1 (55)	133·4 (36)	134·5 ±·22 (191)	136·1±·42 (49)
H′	137·6 (97)	139·1 (51)	138·7 (35)	138·2 ±·24 (183)	139·5±·38 (49)
G′H	74·3 (99)	75·6 (56)	75·0 (36)	74·8 ±·21 (191)	76·1±·39 (46)
J	139·4 (96)	141·1 (53)	137·9 (36)	139·6 ±·32 (185)	140·5±·56 (47)
NH	53·4 (99)	54·6 (56)	53·5 (37)	53·75±·15 (192)	52·4±·26 (48)
100 *B/L*	71·3 (100)	71·6 (54)	70·9 (36)	71·3 ±·15 (190)	71·8±·30 (49)
100 *NB/NH*	43·6 (98)	42·2 (56)	42·9 (37)	43·1 ±·19 (191)	44·3±·36 (48)

* Some of the means in this column differ slightly from the values given in the paper in the *Annals of Eugenics* cited, as the latter were found from distributions instead of by direct addition.

† The probable errors in this column were found by using the standard deviations (given in Table I) for 40 skulls of the same series.

The Eskimo population of Greenland may not be quite as racially homogeneous as Fürst and Hansen supposed, but for practical purposes there can be little harm in combining the three series to form a single sample representing it. Some of the pooled means are given in Table III and they may be compared with those for Hrdlička's Greenland series in the last column of the table. It must be remembered in this case that comparison is being made between measurements taken by different people. Differences greater than three times their probable errors are only found in the case of *B* ($\Delta/(\text{p.e. } \Delta) = 3{\cdot}4$) and *NH* (4·5). There is reason to believe that the latter difference is occasioned by the fact that the nasal height was not determined in precisely the same way for the two series. Hrdlička's means for all the absolute measurements in the table, except *C* and *NH*, are greater than those for Fürst and Hansen's sample. As the facial height (*G′H*) is greater it would be anticipated that the nasal height would also be greater, but actually it is significantly the lesser. As the major calvarial chords (*L*, *B* and *H′*) are greater for Hrdlička's than for Fürst and Hansen's series, the cranial capacity (*C*) would also be expected to be greater for the former, but the difference observed is of the opposite sign, though not significant. The divergences observed thus appear to be partly due to slight differences in the technique of measurement, and the size difference may be partly due to a difference in the process of sexing the skulls. In spite of these blemishes the two series of means are in close agreement. The coefficient of racial likeness between them can be computed for 16 characters

and a crude value of $\cdot 94 \pm \cdot 24$ is obtained. The highest α is for *NH* (8·2) and the next highest for *B* (5·0). It is at least possible that the coefficient differs significantly from zero merely on account of the personal equations of the measurers.

Under the circumstances, it was felt that it would be best to use the pooled means of the two series of Greenland Eskimo skulls for comparative purposes. These are given in column 3 of Table IV and the other means there are for the six series derived from Hrdlička's data, in ways described above, and finally adopted for purposes of comparing different Eskimo types with one another and with non-Eskimo types. Before treating all the Greenland Eskimo skulls as a single sample, however, it was thought advisable to make comparisons between the two sub-samples and the six series relating to Eskimo populations outside Greenland. Crude and reduced coefficients of racial likeness* between these six and the Greenland series measured by (*a*) Hrdlička, (*b*) Fürst and Hansen, and (*c*) Hrdlička and Fürst and Hansen (pooled) are given in Table V. All these coefficients differ from zero with marked significance. The first of the three series is by far the smallest and it gives the lowest crude values in all cases and values markedly lower than the others in five of the six comparisons. Series (*b*) gives intermediate values of the crude coefficients in the case of four out of the six triads, and values only slightly in excess of those for series (*c*) in the other two cases. Corresponding reduced coefficients show a much closer approach to equality, but for five of the six triads the lowest values are with series *a*, while for all six the pooled series gives intermediate values. This last relation suggests that the process of reducing the coefficients is effective, as it appears to give a measure of resemblance independent of the sizes of the samples. If the Greenland series measured by Hrdlička showed the lowest reduced coefficient in the case of comparisons with all six of the other series measured by him, this might be attributed to a difference between his technique of measurement and that employed by Fürst and Hansen. But there is one exception, in the comparisons with the "Old Igloos" series, and another explanation of the results obtained may be suggested. Nothing is known about the origin of the Greenland skulls measured by Hrdlička. If they did not form a true random sample from the total population of the country, but one biased in such a way that it bears a slightly closer resemblance to modern Western Eskimo types than this total population considered as a whole does, then its slightly lower reduced coefficients with five of the six series would be expected. It is shown below that these five are closely related to one another (see Fig. 1), and that the Greenland type does not belong to the same group though it is attached to it. The "Old Igloos" series also diverges from the Western group in the same direction as, but to a greater extent than, the Greenland series. On the hypothesis considered, Hrdlička's Eskimo series would thus be expected to be rather farther removed from the "Old Igloos" series than Fürst and Hansen's Greenland series —supposed to be a random sample from the total population of the country—is

* These coefficients are defined on pp. 100–102 of this volume of *Biometrika*.

TABLE IV

*Mean Measurements of the Male Series of Eskimo Skulls finally adopted**

	"Old Igloos": Point Barrow	Greenland‡	North-Western	Central	Nunivak Island	Point Hope	Western
C	—	1525·8 (217)	1434·6 (45)	1557·6 (17)	1504 (46)	1474 (126)	1473·1 (200)
L	192·5 (27)	188·7 (241)	187·8 (133)	189·3 (43)	188·1 (46)	184·0 (131)	183·9 (229)
B	133·0 (27)	134·7 (240)	137·5 (133)	140·2 (43)	140·9 (46)	138·6 (131)	142·1 (229)
H′	140·2 (27)	138·5 (232)	137·9 (127)	139·0 (43)	136·9 (46)	139·0 (128)	136·6 (220)
LB	107·0 (27)	105·7 (234)	106·0 (127)	106·4 (42)	105·5 (46)	104·9 (128)	103·7 (218)
GL	104·5 (20)	104·6 (229)	104·8 (96)	104·9 (37)	106·5 (42)	103·1 (105)	104·2 (188)
G′H	77·1 (24)†	75·1 (237)	78·3 (97)	76·9 (39)	78·3 (43)	75·2 (118)	78·3 (203)
J	141·6 (26)	139·7 (227)	142·5 (112)	143·6 (42)	143·2 (45)	143·1 (124)	142·0 (221)
NH	54·5 (27)	53·5 (240)	54·6 (119)	54·0 (43)	53·5 (44)	53·6 (126)	54·3 (223)
NB	23·7 (27)	23·0 (239)	23·6 (119)	23·0 (43)	23·5 (44)	23·9 (126)	24·5 (223)
O_2	36·2 (25)	36·3 (235)	36·2 (115)	36·2 (42)	35·9 (42)	36·3 (118)	36·8 (220)
O_1'	39·7 (25)	39·9 (47)	40·4 (115)	40·5 (42)	40·2 (42)	40·3 (118)	40·3 (220)
100 *B*/*L*	69·1 (27)	71·4 (239)	73·2 (133)	74·0 (43)	75·0 (46)	75·3 (131)	77·2 (229)
100 *H′*/*L*	{72·8 (27)}	{73·4 (232)}	{73·4 (127)}	{73·4 (43)}	{72·8 (46)}	{75·5 (128)}	{74·3 (220)}
100 *B*/*H′*	{94·9 (27)}	{97·3 (232)}	{99·7 (127)}	{100·9 (43)}	{102·9 (46)}	{99·7 (128)}	{104·2 (220)}
100 *NB*/*NH*	43·6 (27)	43·4 (238)	43·2 (119)	42·7 (43)	43·8 (44)	44·6 (126)	45·0 (223)
100 O_2/O_1'	91·3 (25)	91·4 (47)	89·6 (115)	89·2 (42)	89·2 (42)	90·1 (118)	91·3 (220)
N∠	{66°·9 (20)}	{68°·1 (229)}	{67°·3 (96)}	{67°·5 (37)}	{68°·8 (42)}	{67°·4 (105)}	{68°·0 (188)}
A∠	{70°·3 (20)}	{69°·9 (229)}	{69°·2 (96)}	{69°·8 (37)}	{67°·9 (42)}	{70°·2 (105)}	{67°·7 (188)}
B∠	{42°·8 (20)}	{42°·0 (229)}	{43°·5 (96)}	{42°·7 (37)}	{43°·3 (42)}	{42°·4 (105)}	{44°·3 (188)}

* The indices and angles in curled brackets were derived from the means of the chords involved, instead of from values for individual skulls.

† Given in error as 261 by Hrdlička. There are 27 skulls in the series, the mean *J* is given for 26 and the mean 100 *G′H*/*J* for 24: hence the mean *G′H* given must be for 24 or 25 skulls.

‡ The Greenland means are the pooled values obtained from Fürst and Hansen's and Hrdlička's series, and all the other series in the table were measured by Hrdlička.

TABLE V

*Coefficients of Racial Likeness between Male Greenland Eskimo Series and other Eskimo Series measured by Hrdlička**

Greenland Eskimo series measured by	"Old Igloos": Point Barrow (25·7)	North-Western (114·1)	Central (40·2)	Point Hope (123·7)	Nunivak Island (44·6)	Western (215·9)
	Crude coefficients					
Hrdlička (*H.*) (47·0)*	2·99 ± ·25 (15)	4·63 ± ·24 (16)	3·51 ± ·24 (16)	10·06 ± ·24 (16)	7·33 ± ·24 (16)	29·22 ± ·24 (16)
Fürst and Hansen (*F.*) (186·6)*	3·20 ± ·25 (15)	12·38 ± ·24 (16)	8·58 ± ·24 (16)	26·30 ± ·24 (16)	16·14 ± ·24 (16)	86·30 ± ·24 (16)
H. and *F.* (pooled) (233·2)*	3·28 ± ·25 (15)	12·63 ± ·24 (16)	8·19 ± ·24 (16)	27·73 ± ·24 (16)	15·85 ± ·24 (16)	94·86 ± ·24 (16)
	Reduced coefficients					
Hrdlička (*H.*)	8·98 ± ·73	6·95 ± ·36	8·10 ± ·55	14·76 ± ·35	16·01 ± ·52	37·86 ± ·31
Fürst and Hansen (*F.*)	7·09 ± ·54	8·75 ± ·17	12·97 ± ·36	17·67 ± ·16	22·41 ± ·33	43·07 ± ·12
H. and *F.* (pooled)	7·10 ± ·53	8·25 ± ·16	11·95 ± ·35	17·16 ± ·15	21·16 ± ·32	42·31 ± ·11

* All the coefficients in this table are based on the same 16 characters, viz. *C*, *L*, *B*, *H'*, *LB*, *J*, *G'H*, *NH*, *NB*, O_2, 100 *B/L*, 100 *H'/L*, 100 *B/H'*, 100 *NB/NH*, *N*∠ and *A*∠, except those with the "Old Igloos" series for which the capacity (*C*) is not available. In the case of the Greenland Eskimo series the mean numbers of skulls given are for all 16 characters. These are changed slightly when *C* is omitted, the means for the remaining 15 being *H.* 47·3, *F.* 187·3, *H.* and *F.* 234·9.

from the "Old Igloos" series. The reduced coefficients show these relationships, but it is clear that the hypothesis cannot be justified rigorously. Only the pooled Greenland series was used in later comparisons.

Seven series were thus finally adopted: the means for these are given in Table IV and the reduced coefficients of racial likeness between them in Table VI. Fig. 1 shows the localities from which the material was obtained and the connections provided by the reduced coefficients less than 10. Three of the Alaskan and the Central series are all closely connected with one another. The Western diverges from this central group in one direction and the Greenland series diverges from it in a different direction. So far there is a general agreement between the

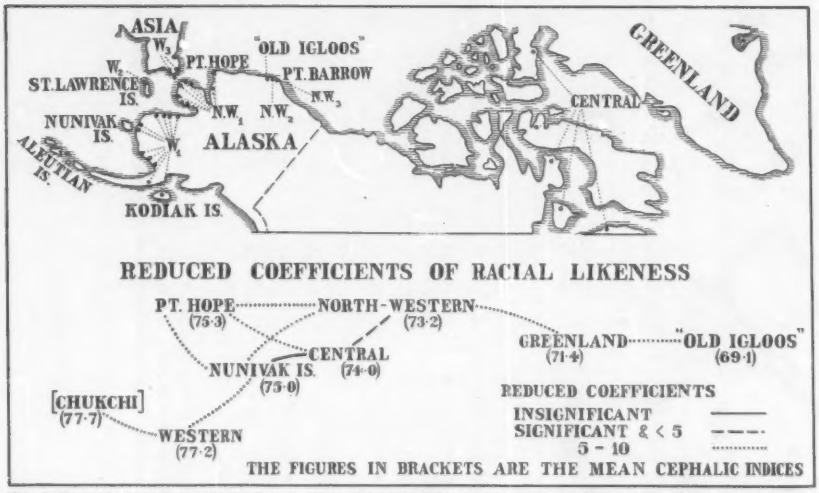


FIG 1 THE PLACES OF ORIGIN & RELATIONSHIP OF MALE SERIES OF ESKIMO SKULLS.

relationships found and the geographical positions of the populations represented, but this is not maintained in detail. The Greenland series, for example, is nearest geographically to the Central, but it bears a closer resemblance to the North-Western Alaskan than to the Central series. The extremely close resemblance of the Nunivak Island and Central Eskimo types is again unexpected. The only coefficient which differs from zero by less than three times its probable error is found in this case, but the two series show distinctly different relationships when compared with the others. It has been found in the case of other material that the fact that two series cannot be clearly differentiated when compared directly does not preclude the possibility that they will be distinguished by other comparisons. The relationships of the "Old Igloos" series have not been considered yet. This was obtained from a site in Alaska within a few miles of some of the others, but it differs from all the other Eskimo series in being older than they are, and it is also

TABLE VI

*Reduced Coefficients of Racial Likeness between Male Eskimo Cranial Series**

		Greenland	"Old Igloos": Point Barrow	North-Western	Central	Point Hope	Nunivak Island	Western
	$\bar{n}$	213·1	25·6	114·2	40·4	123·1	44·3	216·3
Greenland	213·1	—	6·10 ± ·49 (17)	7·76 ± ·15 (18)	11·16 ± ·33 (18)	15·88 ± ·14 (18)	19·44 ± ·31 (18)	39·55 ± ·10 (18)
"Old Igloos": Point Barrow	25·6	6·10 ± ·49 (17)	—	16·23 ± ·55 (17)	24·07 ± ·73 (17)	32·91 ± ·55 (17)	38·44 ± ·71 (17)	67·57 ± ·50 (17)
North-Western	114·2	7·76 ± ·15 (18)	16·23 ± ·55 (17)	—	2·61 ± ·38 (18)	7·16 ± ·19 (18)	5·49 ± ·35 (18)	18·47 ± ·15 (18)
Central	40·4	11·16 ± ·33 (18)	24·07 ± ·73 (17)	2·61 ± ·38 (18)	—	7·00 ± ·37 (18)	1·52 ± ·53 (18)	17·28 ± ·33 (18)
Point Hope	123·1	15·88 ± ·14 (18)	32·91 ± ·55 (17)	7·16 ± ·19 (18)	7·00 ± ·37 (18)	—	8·92 ± ·35 (18)	10·93 ± ·14 (18)
Nunivak Island	44·3	19·44 ± ·31 (18)	38·44 ± ·71 (17)	5·49 ± ·35 (18)	1·52 ± ·53 (18)	8·92 ± ·35 (18)	—	7·11 ± ·31 (18)
Western	216·3	39·55 ± ·10 (18)	67·57 ± ·50 (17)	18·47 ± ·15 (18)	17·28 ± ·33 (18)	10·93 ± ·14 (18)	7·11 ± ·31 (18)	—

* The means for the 18 characters on which these coefficients are based are given in Table IV, all the characters there being used except *GL* and $B\angle$. The coefficients with the "Old Igloos" series are for 17 characters, as there is no mean capacity (*C*) available for it. The mean numbers of skulls for the characters used ($\bar{n}$'s) are for all 18 in the case of all the series except the "Old Igloos". In the comparison of this with the six others their $\bar{n}$'s are slightly different from the values given for all 18 characters.

the smallest series used. The type shows no affinities to any of the modern ones found in the same area and it only shows one close connection, which is with the Greenland series.

The material available shows that there are a number of existing Eskimo populations of types which can be clearly distinguished from one another, and there is evidence of one type which appears to be extinct to-day. The fact that no close correlation is found between geographical position and resemblance as measured by the reduced coefficient of racial likeness may be due to migrations of the peoples represented. Evidence of other extinct Eskimo populations will probably be needed in order to throw light on the origin of the present-day varieties, but comparison with non-Eskimo material should also aid a solution of this problem.

6. *Comparisons of Single Characters.* The means of the seven Eskimo series finally adopted for purposes of racial comparison are given in Table IV. They relate to 20 characters and the coefficients of racial likeness were computed for 18 of these. In the process of computation a value, α, is obtained for each character in each comparison; this is approximately the square of the ratio of the difference between two means to its standard error. For the seven series there are $7 \times 6/2 = 21$ comparisons for each character, except in the case of the capacity (C), for which the total is 15, as one mean is missing. We may decide, quite arbitrarily, to consider that an α indicates a significant difference if it is greater than 10. If samples were drawn from two populations which actually had identical means for a particular character, then an α greater than 10 would only be expected to occur once in about 625 trials. The numbers of α's greater than 10 in the 21 comparisons will give estimates of the relative degrees to which the coefficients are determined by different characters. As is usually found in such comparisons, there are marked distinctions between the characters when examined in this way, some being practically constant for all the series and others showing significant differences between most pairs of them.

The three orbital measurements and the nasal angle show no α's at all greater than 10. The nasal height is almost as constant, as it only shows one value greater than the limit chosen and this is only slightly above the limit: the α is 11·30 for the Greenland and North-Western series. A second group of characters may be distinguished by the fact that they only show significant differences in some of the comparisons between the Western series on the one hand, and the other six series on the other. These are the nasal index (3 α's > 10), the basio-bregmatic height (3 α's > 10) and the chord from nasion to basion (LB: 4 α's > 10). It can be seen from Table IV that the Western series has the highest mean for $100\,NB/NH$ and the lowest means for H' and LB. Two other characters may be added to this group. The alveolar angle ($A\angle$) show 7 of the 21 α's greater than 10, and 5 of these —including the only two α's indicating markedly significant differences—are for

comparisons between the Western and the other series: the Western has the lowest mean. The nasal breadth shows 5 α's greater than 10, and 4 of these are for comparisons between the Western and the other series: the Western has the highest mean. This disposes of 10 of the 18 characters: 5 of these 10 may be supposed practically constant for all the Eskimo types, and the other 5 characters are practically constant for 6 series, but they distinguish these 6 from the Western Eskimo series.

The remaining 8 characters serve other purposes. The bizygomatic breadth (J) only shows 5 of its 21 α's significant, though all these indicate clear divergence, and they are all for comparisons between the Greenland Eskimo and the other series: the Greenland mean is not distinguished from the "Old Igloos", but it is from all the others. The height-length index again distinguishes one series from all the others, but in this case it is the Point Hope: all its α's for the character indicate clear significance, and there is only one other α greater than 10 (viz. 10·52). The significant differences are more erratic in the case of the upper facial height ($G'H$: 6 α's out of 21 greater than 10) and of the capacity (C: 5 α's out of 15 greater than 10). The remaining 4 characters are distinguished from all the others by the fact that they show more significant than insignificant differences. In each one of these cases there are 21 comparisons and the numbers of α's greater than 10 are 13 for L, 14 for B, 15 for $100\,B/H'$ and 17 for $100\,B/L$.

In the comparison of a group of series representing closely related populations, it is commonly found that the major calvarial chords and the indices derived from them show a larger percentage of significant differences than any other characters, and the cephalic index generally distinguishes the types more effectively, on the average, than the calvarial length or breadth from which it is derived. In Table IV the series are arranged in order of their mean cephalic indices: L, B and $100\,B/H'$ give very similar orders to this, but the same is not true for any other character. By considering the characters singly and then attempting to combine the evidence of each, it does not seem to be possible to construct any clear picture of the situation. Different characters suggest different conclusions and the advantage of using a generalized criterion, such as the coefficient of racial likeness, is evident.

The coefficients (Fig. 1) suggest that four of the Eskimo series represent populations which are all closely related to one another. The Greenland Eskimos diverge from this central group in one direction, and the type known from the "Old Igloos" skulls diverges in the same direction, but to a greater extent: the Western Eskimos diverge from the central group in another direction. This arrangement is suggested by the calvarial breadth (B) and two indices ($100\,B/L$ and $100\,B/H'$) which involve this measurement. The same arrangement is not suggested by any other character for which means are given in Table IV, but it is by another index involving B, viz. the "cranio-facial" ($100\,J/B$). Means for this, computed not from individual measurements but from the means of the two

chords, are: "Old Igloos" 106·5 (26), Greenland 103·7 (227), North-Western 103·6 (112), Central 102·4 (42), Nunivak Island 101·6 (45), Point Hope 103·2 (124) and Western 99·9 (221). Dr von Bonin has given* means of this index for 49 male cranial series and the highest value in his list is 104·9 for Loyalty Islanders.

The mean calvarial length of the "Old Igloos" series (192·5) is almost as great as the largest recorded for any male series of skulls; the nasal breadth of 23·0 (Greenland and Central series) is very close to the extreme found for all races, and the nasal index of 42·7 (Central) appears to be the lowest as yet recorded. These, however, are not characters which distinguish the "Old Igloos" and Greenland from all the other Eskimo types.

It may be noted that the Eskimo skull also appears to be quite extreme among modern races of man in having the "flattest" facial skeleton, though its nasal bridge is not peculiarly flat.† It has also been shown that its malar bones are extremely large and that an index expressing their vertical arcs as percentages of their horizontal arcs makes a clear distinction between the Eskimo and all other races for which data are available.‡ These measurements have only been given for a single series of Eskimo skulls—viz. one made up principally by specimens from Greenland—and their means for the series measured by Dr Hrdlička should be of particular interest. Other features which cannot be estimated from any measurements available, such as the median sagittal crest, also demonstrate that the Eskimo type is peculiarly specialized.

7. *Comparisons between Eskimo and Asiatic Series.* As a preliminary to any discussion of the "origin" of the Eskimo, it is clear that comparisons must be made between the different varieties found and series representing other races. The type is certainly one of the most specialized known, and the fact that several distinct varieties of it are found should aid the solution of problems concerning its relationships. In spite of the striking resemblance of the Chancelade to modern Eskimo skulls, there is no race known to have existed in Europe since palaeolithic times which is closely similar to that of the northern people. It is to be expected, however, that close affinities will be found with certain American and Asiatic races. It is hoped that the results of statistical comparisons between the Eskimo and North American cranial material will be presented later,§ and only comparisons with Asiatic material are considered here.

Coefficients of racial likeness for all pairs of 26 male Asiatic series have been published.‖ Their full comparison by the same method with the seven Eskimo

* *Biometrika*, XXVIII (1936), p. 133.

† See T. L. Woo and G. M. Morant, "A Biometric Study of the 'Flatness' of the Facial Skeleton in Man", *ibid.* XXVI (1934), pp. 196–250.

‡ See T. L. Woo, "A Biometric Study of the Human Malar Bone", *ibid.* XXIX (1937), pp. 113–23.

§ In a paper by Dr von Bonin and the writer which is nearly completed.

‖ T. L. Woo and G. M. Morant, "A Preliminary Classification of Asiatic Races based on Cranial Measurements", *Biometrika*, XXIV (1932), pp. 108–34. The Tibetan *B* series of 15 skulls was omitted because it is too short for the purpose in view.

series would necessitate the computation of 182 coefficients. But it has been decided that for purposes of classification the lower orders of reduced coefficients only should be used, and it is to be anticipated that the vast majority of these 182 will be of higher orders which can be neglected. The method described below makes it possible to decide, from a comparison of the means for a few characters, whether the sets of means for two series may possibly give a reduced coefficient less than a particular value (19), or whether it will be safe to assume that the coefficient will be greater than this limit. If the simple test indicates the second of these conclusions, then there is no need to calculate the coefficient, as it will not be needed in the classification. The method has been used in the comparison of other groups of series, and it makes it unnecessary to carry out a large amount of computation.

In classifying the 26 Asiatic series, all reduced coefficients less than 19 were neglected. It was found for the total 325 ($=26\times25/2$) comparisons that the calvarial length, breadth and height and the three indices derived from these chords gave numbers of significant α's larger than, or almost as large as, the numbers given by any other of the 31 characters used. The values of the coefficients are evidently determined in large part by these six measurements, though others also play important rôles. The maximum differences between the means found in the case of the 54 comparisons giving reduced coefficients less than 19 are:

L	B	H'	$100\,B/L$	$100\,H'/L$	$100\,B/H'$
6·7 mm.	6·1 mm.	6·3 mm.	5·4	3·4	6·5

These values are much less, of course, than the corresponding maximum differences which would be found in the case of all possible comparisons between pairs of the 26 Asiatic series. If any one of these series could be compared with a new Asiatic series, and if any one of the differences of the means for the six characters were found to be greater than the value for the character given above, then it is unlikely that the reduced coefficient found in this case would be less than 19. Under the same circumstances, it is still less likely that one of the Asiatic and a non-Asiatic series would give a reduced coefficient less than 19. These considerations can be used to select those pairs of series in new comparisons which will be the only ones likely to provide reduced coefficients less than the limit which has been arbitrarily chosen. The ranges of the differences actually used for this purpose were those above with the addition of ·1 to each, viz. L 6·8 mm., B 6·2 mm., H' 6·4 mm., $100\,B/L$ 5·5, $100\,H'/L$ 3·5 and $100\,B/H'$ 6·6.

Comparisons of means restricted to these six characters were first made between each of the seven Eskimo series, on the one hand, and each of the 26 Asiatic series, on the other. If for a particular pair the difference found between the means was found to be in excess of the limit fixed in the case of any one or more of the characters then no calculation was carried out, as it may be presumed that all such pairs would give reduced coefficients of racial likeness greater than

19. This speedy test made it unnecessary to calculate 161 of the total 182 coefficients possible. The remaining 21* might give connections of the order finally considered, but the majority of them would still be expected to be of a higher order. In each of these cases the α's were first calculated for characters which showed most significant differences, and it was generally possible to see from a few of these that the reduced coefficient must exceed 19. It was only necessary to calculate two coefficients in full, and one of these was found to exceed 19 while the other is:

Western Eskimo ($\overline{n} = 220{\cdot}8$) and Chukchi (34·1)—reduced coefficient $= 7{\cdot}05 \pm {\cdot}45$ for 13 characters.

In this comparison the nasal index ($\alpha = 14{\cdot}0$) is the only character which gives an α greater than 10. The Chukchi series (measured by Fridolin†) represents a people, inhabiting the extreme north-east of Asia, generally supposed to have close physical affinities to the Eskimos. It only showed one reduced coefficient less than 19 with the other Asiatic series, viz. that of $18{\cdot}27 \pm {\cdot}65$ with the Prehistoric Chinese. The modern Chinese can thus be linked to the Greenland Eskimo type by a number of intermediate types, the sequence being: Modern Chinese—Prehistoric Chinese—Chukchi—Western Eskimo—Central Eskimo—Greenland Eskimo.

* The test also allows four comparisons between the Tibetan *B* and the Eskimo series, but all the reduced coefficients were found to be greater than 19.

† The calvarial height given for the Chukchi skulls is the vertical from the basion (*H*) in place of the more usual basio-bregmatic (*H'*). One mm. was subtracted from the mean *H* to give an approximation to *H'*, as average differences very close to this have been found for all the longest series for which both heights have been given.

ON THE z-TEST IN RANDOMIZED BLOCKS AND LATIN SQUARES

By B. L. WELCH, Ph.D.

CONTENTS

SECT.		PAGE
1.	Introduction	21
2.	Randomized Blocks	22
3.	Normal Theory and Randomization Compared	24
4.	A Particular Example (Randomized Blocks)	28
5.	Further Examples (Randomized Blocks)	29
6.	Latin Squares (Normal Theory)	31
7.	Latin Squares (Randomization Theory)	32
8.	The 4×4, 5×5 and 6×6 Squares	41
9.	Examples (Latin Squares)	44
10.	Summary and Conclusions	47
APPENDIX		
A.	Derivation of Expectations in Equation (48)	48
B.	Confirmation of Equation (49)	52

1. *Introduction.* The validity of Fisher's z-test in the practical situations in which it has been applied has been the subject of much discussion.* In general, the mathematical distribution of z follows from the assumption that the sample observations x_i $(i=1, 2, \ldots, n)$ can be written

$$x_i = c_{i1}\theta_1 + c_{i2}\theta_2 + c_{i3}\theta_3 \ldots c_{ir}\theta_r + \eta_i, \qquad \ldots\ldots(1)$$

where the c's are known numbers, the θ's are $r<n$ unknown parameters, and the η's are normally and independently distributed about zero with standard deviations proportional to known numbers. Results following from such a starting-point may be termed results from normal theory. In practice we may not wish to make all the above assumptions regarding the η's, and to a certain extent it can be shown that, not doing so, we can still use the tests based upon them. Of especial interest are the cases of experimentation into which randomization enters as part of the structure. R. A. Fisher has pointed out† that, in any such case, it is possible to carry through arithmetical calculations, from which the hypothesis under test may be judged, without making any assumptions whatever. These calculations are lengthy. One can, however, consider only certain

* For a bibliography of the subject see a paper by T. Eden and F. Yates entitled "On the Validity of Fisher's z-test when Applied to an Actual Example of Non-Normal Data", *J. agric. Sci.* XXIII (1933), pp. 6–16. Other references are given later.

† See, for instance, *The Design of Experiments*, Oliver and Boyd (1935), p. 51.

aspects of them, which are sufficient to give useful comparisons with the results from normal theory. This I have done in the present paper.

2. *Randomized Blocks.* The first situation to be discussed will be that of Randomized Blocks. Here the treatments under test are represented once and once only in each of a number of blocks. The yields given by an experiment may be denoted by $x_{i(k)}$, where $i\,(=1, 2, \ldots, n)$ are the blocks and $k\,(=1, 2, \ldots, s)$ the treatments. The usual procedure employed to test whether the treatments can be regarded as equivalent is to perform on the yields the analysis of Table I and

TABLE I

Analysis of Variance for Randomized Blocks

Source	Degrees of Freedom	Sum of Squares	Mean Square
Between Treatments	$f_1=(s-1)$	S_1	$v_1=S_1/f_1$
Between Blocks	$f_2=(n-1)$	S_2	$v_2=S_2/f_2$
Residual	$f_0=(n-1)\,(s-1)$	S_0	$v_0=S_0/f_0$
Total	$(ns-1)$	S	

to calculate the criterion $z=\frac{1}{2}\log_e(v_1/v_0)$. This criterion is then referred to a certain theoretical distribution, namely the distribution of z obtained by assuming that

$$x_{i(k)}=A_i+\eta_{ik}, \qquad \ldots\ldots(2)$$

where the A's are unknown block means and the η's are all normally and independently distributed about zero, with the same unknown standard deviation.

To investigate the meaning and extent of these assumptions it is necessary to consider further details of the experimental arrangement, and the exact manner in which the hypothesis, that the treatments are equivalent, is usually formulated. For convenience let the plots in each block be numbered $j=1, 2, \ldots, s$. Then the yield which the kth treatment would give, if applied under the experimental conditions, on the jth plot of the ith block may be denoted by $x_{ij(k)}$. For any plot (i, j), of course, only one of the quantities $x_{ij(k)}$ is real, viz. the one for the treatment k which is actually used on that plot in the experiment. The other $x_{ij(k)}$ are hypothetical, based on the conception of what might happen if the experiment could be repeated under the same essential conditions, using in turn every treatment on the plot (i, j). The hypothesis that the experiment is to test must, for statistical purposes, be expressed as a relation holding in some hypothetical population. In this case the population consists of the values $x_{ij(k)}$.* In the

* For a further discussion of this manner of defining our statistical population see the concluding section of the paper.

literature the hypothesis has been formulated in terms of these x's, in two distinct ways.

(i) R. A. Fisher has considered the hypothesis that the treatments would give equivalent results on *every individual plot*, i.e. that $x_{i(k)}$ would be the same for all k.

(ii) J. Neyman has suggested that we should allow the possibility that the treatments would affect individual plots differentially, and should consider the hypothesis that the average yield of each treatment, if applied over the *whole experimental field*, would be the same. This means that $x_{\cdot\cdot(k)}$ would be the same for all k, where $x_{\cdot\cdot(k)}$ is the mean of $x_{i(k)}$ over all (i, j).

The first of these "null"* hypotheses is the one which will be considered here. We may express it by the following equation:

$$x_{ij(k)} = x_{ij} \qquad (k = 1, 2, \ldots, s). \qquad \ldots\ldots(3)$$

The other "null" hypothesis was discussed in a paper given recently by Dr Neyman to the Royal Statistical Society,† in relation to the same problem that I am concerned with here. He, too, investigated the influence on the z-test, of the fact that the assumptions of normality and independence in equation (2) are not exactly satisfied. He came to certain conclusions and expressed the opinion that further investigation was desirable. The results that I obtain in the present paper may, I think, profitably be compared with his. I should emphasize, however, that the "null" hypothesis which I am using is that of equation (3) and that the situations are therefore not exactly the same. However, as Dr Neyman points out in commenting on the discussion after his paper, his results are applicable to the "null" hypothesis of (3), if some of the quantities in his equations are given certain values. On the other hand the methods I adopt here are not applicable to his more general "null" hypothesis.

We must now refer to the essential point of the arrangement of Randomized Block experiments. This is as follows. In every block the s treatments are assigned entirely at random to the s plots available for them. This means that, if the hypothesis of equivalent treatments is true (i.e. if (3) is satisfied), the yields $x_{i(k)}$ $(k = 1, 2, \ldots, s)$, given by the experiment will be a random arrangement of x_{ij} $(j = 1, 2, \ldots, s)$. For instance, in Fig. 1 (a) is given a possible set of yields x_{ij} for a field consisting of four blocks with three plots each. Fig. 1 (b) shows one possible way in which the treatments may be arranged on this field. In the first block, treatments 1, 2 and 3 are on plots 2, 1 and 3, respectively; this is only one of 3! possible arrangements. Similarly, in the other three blocks we have illustrated one of the 3! possible arrangements. Hence, taken as a whole, Fig. 1 (b) represents one of $(3!)^4$ possible arrangements.

* The term "null hypothesis" is used in the literature to denote the hypothesis that the treatments are equivalent.

† "Statistical Problems in Agricultural Experimentation", *J. Roy. statist. Soc.* Suppl. II No. 2 (1935), pp. 107–180.

The method of randomization gives all the possible arrangements an equal chance of occurring. Corresponding to each arrangement there will be a different reclassification of the yields by block and treatment. Fig. 1 (*c*), for instance, shows the reclassification of the yields of Fig. 1 (*a*) when the arrangement Fig. 1 (*b*) is applied to the field. From the reclassified yields the analysis of variance of Table I can be worked out and the value of z computed. The $(3!)^4$ arrangements will each lead to a value of z, and the z of the experiment may be regarded as randomly selected from this distribution of values. The question whether the theoretical z-distribution of normal theory gives a valid test of the hypothesis of treatment equivalence, therefore, involves a comparison of the theoretical distribution of z with the distribution which would be obtained by taking all the possible arrangements.*

		Numbers of Plots (j)		
		1	2	3
Blocks (i)	1	16	12	10
	2	26	22	24
	3	14	11	6
	4	31	37	40

Fig. 1 (*a*). Example of possible yields x_{ij} on 4 block by 3 plot field.

		Numbers of Plots (j)		
		1	2	3
Blocks (i)	1	2	1	3
	2	3	1	2
	3	1	3	2
	4	2	3	1

Fig. 1 (*b*). Possible arrangement of treatments k (=1, 2, 3) on the field of Fig. 1 (*a*).

		Treatments (k)		
		1	2	3
Blocks (i)	1	12	16	10
	2	22	24	26
	3	14	6	11
	4	40	31	37

Fig. 1 (*c*). Reclassification of yields $x_{i(k)}$ obtained by applying the arrangement Fig. 1 (*b*) to the field Fig. 1 (*a*).

3. *Normal Theory and Randomization Compared.* One approach to the comparison of the z-distribution from normal theory with that from randomization is to take separately the mean squares v_0 and v_1 of which z is a function. In normal theory v_0 and v_1 are independently distributed and their mean values are both σ^2, where σ is the standard deviation of a single η in equation (2). From randomization it is found that v_0 and v_1 have equal expectations. They are not, however, independent, since the sum $(S_0 + S_1)$ is constant, being the total sum of squares within blocks and therefore not dependent on the manner in which the treatments are assigned within blocks. The parallelism between the two theories also breaks down if we consider the variances of v_0 and v_1.

In the normal theory v_0 and v_1 are distributed as $(\chi_0^2\sigma^2)/f_0$ and $(\chi_1^2\sigma^2)/f_1$ respectively, where χ_0^2 and χ_1^2 are independent χ^2's with f_0 and f_1 degrees of freedom. The variances of v_0 and v_1 are therefore $2\sigma^4/(n-1)(s-1)$ and $2\sigma^4/(s-1)$, i.e. are in the ratio of $1:(n-1)$. From randomization, since $(S_0 + S_1)$ is constant, the variance of S_0 must be the same as that of S_1. v_0 and v_1 therefore have variances

* For brevity, any frequency constants calculated over all the possible random arrangements will be termed constants calculated from randomization. It should be emphasized that we are only considering what happens if the hypothesis of equation (3) is really true.

proportional to $1/f_0^2$ and $1/f_1^2$, i.e. variances in the ratio $1:(n-1)^2$. With such discrepancies between results from normal theory and from randomization, it would be unsafe to conclude, from the circumstance that the expectations of v_0 and v_1 are in agreement, that the z-distributions will also compare favourably. It is advisable to try to obtain some results for the z-distribution directly,* and this I have done by means of a simple transformation.

It is of interest to note here a sampling experiment which was designed to compare directly the z-distribution from randomization with that from normal theory, for one particular set of data. This practical investigation, carried out by T. Eden and F. Yates,† did show close agreement between the z-distributions obtained in the two ways. We shall return to their example later.

Instead of z, we can equally well use the function of z:

$$U = \frac{S_1}{S_0 + S_1} = (\overline{n-1}\, e^{-2z} + 1)^{-1}, \qquad \text{.....(4)}$$

which increases monotonically with z. Instead of saying that we reject the hypothesis of equivalent treatments when $z > z_0$ (say), we shall now say that we reject when $U > U_0$, where U_0 is related to z_0 by equation (4). If the U distributions from normal theory and from randomization compare favourably, then necessarily the z-distributions will do so also. The convenience of U lies in the fact that in the randomization procedure $(S_0 + S_1)$ is constant, and thus only the variation of S_1 need be considered.

The comparison of the U distributions will be made through the medium of their first two moments. In the normal theory we have

$$U = \frac{(\chi_1^2 \sigma^2)}{(\chi_0^2 \sigma^2 + \chi_1^2 \sigma^2)} = \frac{\chi_1^2}{(\chi_0^2 + \chi_1^2)},$$

where χ_0^2 and χ_1^2 are independently distributed as χ^2 with degrees of freedom $f_0 = (n-1)(s-1)$ and $f_1 = (s-1)$, respectively. It follows that the distribution of U is

$$p(U) = \text{const.} \times U^{(f_1/2)-1} (1-U)^{(f_0/2)-1}. \qquad \text{.....(5)}$$

The moments of this distribution are

$$\begin{cases} \text{mean } U_N = \dfrac{f_1}{f_0 + f_1} = \dfrac{1}{n}, \\ {}_N\mu_2' = \dfrac{f_1(f_1+2)}{(f_0+f_1)(f_0+f_1+2)} = \dfrac{s+1}{n(ns-n+2)}, & \text{.....(6)} \\ \sigma_{U_N}^2 = \dfrac{2(n-1)}{n^2(ns-n+2)}. & \text{.....(7)} \end{cases}$$

* Dr Neyman in the discussion after his paper already referred to, pointed out this advisability of considering the z-distribution directly, when any investigation of the validity of the z-test is being made.

† "On the Validity of Fisher's z-test", *loc. cit.*

The suffix N is here used to denote that the moments are calculated from the normal theory. The suffix R will be used for the moments of U from randomization. The calculation of these moments is made easier by denoting the means of the yields x_{ij} over the blocks by B_i, and the deviations from these means by u_{ij}. Thus

$$x_{ij} = B_i + u_{ij}, \qquad \text{.....(8)}$$

where $\sum_j u_{ij}$ is zero. The experimental yields $x_{i(k)}$ will then be written

$$x_{i(k)} = B_i + y_{i(k)}. \qquad \text{.....(9)}$$

Using the dot notation to indicate that a mean is being taken over all values of the letter replaced by the dot, we have

$$(S_0 + S_1) = \sum_i \sum_k (x_{i(k)} - x_{i(\cdot)})^2 = \sum_i \sum_k y^2_{i(k)},$$

since $x_{i(\cdot)}$ must equal B_i. Also, since $x_{i(k)}$ $(k = 1, 2, \ldots, s)$ are a random arrangement of x_{ij} $(j = 1, 2, \ldots, s)$, $y_{i(k)}$ must be a random arrangement of u_{ij}. Hence

$$(S_0 + S_1) = \sum_i \sum_j u^2_{ij}. \qquad \text{.....(10)}$$

Also
$$S_1 = \sum_k n\,(x_{\cdot(k)} - x_{\cdot(\cdot)})^2 = \sum_k n y^2_{\cdot(k)} = \frac{1}{n} \sum_k \{(\sum_i y_{i(k)})^2\}$$

$$= \frac{1}{n} \sum_k \{\sum_i y^2_{i(k)} + \sum_{i \neq m} y_{i(k)} y_{m(k)}\}^*$$

$$= \frac{1}{n} \sum_i \sum_j u^2_{ij} + \frac{1}{n} \sum_{i \neq m} \sum_k y_{i(k)} y_{m(k)}.$$

Since $\sum_j u_{ij}$ is zero, the expectation of any $y_{i(k)}$ is zero. Further, since the arrangements in different blocks are made independently, the y's in different blocks are independent. Hence, using E to denote expectation, we have

$$E(S_1) = \frac{1}{n} \sum_i \sum_j u^2_{ij}. \qquad \text{.....(11)}$$

Hence
$$\text{mean } U_R = \frac{E(S_1)}{(S_0 + S_1)} = \frac{1}{n}, \qquad \text{.....(12)}$$

which is the same as the mean from normal theory. For the variance of U_R, first consider

$$S_1^2 = \frac{1}{n^2} [\sum_k \{\sum_i y_{i(k)}\}^2]^2,$$

i.e.
$$n^2 S_1^2 = \sum_k \sum_{k'} \sum_i \sum_m \sum_p \sum_q y_{i(k)} y_{m(k)} y_{p(k')} y_{q(k')}.$$

This summation is taken for k and k' over 1, 2, ..., s, and for i, m, p, and q over 1, 2, ..., n. There are thus $n^4 s^2$ terms, but not all of these contribute to the expecta-

* Single summations are over all values of the letter indicated. $\sum_{i \neq m}$ indicates a summation over all values of i and m excluding $i = m$, i.e. the summation includes both $y_{1(k)} y_{2(k)}$ and $y_{2(k)} y_{1(k)}$, the fact that these terms are the same being ignored. This convention is used throughout.

tion of $n^2S_1^2$. For if, in any term, any one of the subscripts i, m, p and q is not equal to some one of the other three, the expectation of that term must be zero. Hence

$$\begin{aligned} n^2E[S_1^2] &= E[\sum_k\sum_{k'}\{\sum_i y_{i(k)}^2 y_{i(k')}^2 + \sum_{i\neq m} y_{i(k)}^2 y_{m(k')}^2 + 2\sum_{i\neq m} y_{i(k)}y_{m(k)}y_{i(k')}y_{m(k')}\}] \\ &= E[\sum_k\sum_{k'}\{(\sum_i y_{i(k)}^2)(\sum_i y_{i(k')}^2) + 2\sum_{i\neq m} y_{i(k)}y_{m(k)}y_{i(k')}y_{m(k')}\}] \\ &= E[\{\sum_k\sum_i y_{i(k)}^2\}^2] + 2\sum_k\sum_{k'}\sum_{i\neq m} E[y_{i(k)}y_{i(k')}]\,E[y_{m(k)}y_{m(k')}] \\ &= (\sum_i\sum_j u_{ij}^2)^2 + 2s\sum_{i\neq m} E[y_{i(k)}^2]\,E[y_{m(k)}^2] \\ &\qquad + 2s(s-1)\sum_{i\neq m} E[y_{i(k)}y_{i(k')}]\,E[y_{m(k)}y_{m(k')}], \qquad \text{.....(13)} \end{aligned}$$

k and k' in the last term standing for any two different treatments.

Now $E[y_{i(k)}^2] = (\sum_j u_{ij}^2)/s$, and for $k \neq k'$ we have

$$E[y_{i(k)}y_{i(k')}] = \frac{\sum_{j\neq j'} u_{ij}u_{ij'}}{s(s-1)} = \frac{-\sum_j u_{ij}^2}{s(s-1)}.$$

Hence (13) gives

$$\begin{aligned} n^2E(S_1^2) &= (\sum_i\sum_j u_{ij}^2)^2 + \frac{2}{s-1}\sum_{i\neq m}(\sum_j u_{ij}^2)(\sum_j u_{mj}^2) \\ &= \frac{(s+1)(\sum_i\sum_j u_{ij}^2)^2 - 2\sum_i(\sum_j u_{ij}^2)^2}{(s-1)}. \end{aligned}$$

Hence from (10), $$E(U_R^2) = \frac{(s+1)-2A}{n^2(s-1)}, \qquad \text{.....(14)}$$

where $$A = \frac{\sum_i(\sum_j u_{ij}^2)^2}{(\sum_i\sum_j u_{ij}^2)^2}. \qquad \text{.....(15)}$$

Since the mean U_R is $1/n$, we get from (14),

$$\sigma_{U_R}^2 = \frac{2(1-A)}{n^2(s-1)}. \qquad \text{.....(16)}$$

As the mean U is the same from normal theory and from randomization, a comparison of the two U distributions can be made, to a first approximation, from the variances of equations (7) and (16). In this discussion attention will be focussed on what has been termed by Neyman and Pearson *the first kind of error*, i.e. the risk of rejecting the hypothesis that the treatments are equivalent when it is actually true, as distinct from *the second kind of error* which occurs when we fail to detect differentiation where it really exists. Suppose we wish the risk of the first kind of error to be ϵ. Let U_0 be the value from normal theory such that $P(U_N > U_0) = \epsilon$. Then the rule adopted to test whether the experimental results are consistent with there being no real differences in treatments, is to reject this

hypothesis if the U of the experiment is greater than U_0. The real chance of the first kind of error should not, however, be drawn from the U_N distribution but from the U_R distribution. Therefore we need to know how $P(U_R > U_0)$ compares with ϵ. Roughly, we may say that if $\sigma_{U_R} < \sigma_{U_N}$ the chance of $U_R > U_0$ will be less than ϵ: on the other hand if $\sigma_{U_R} > \sigma_{U_N}$ the chance of $U_R > U_0$ will be greater than ϵ.

From equation (16) it is seen that apart from n and s, the comparison of σ_{U_N} and σ_{U_R} involves only the single function A of the plot yields. This function depends on the relative sizes of $(\sum_j u_{ij}^2)$ in the different blocks, i.e. on the relative sizes of the observed variances within the blocks. Now the minimum value A can have is when $(\sum_j u_{ij}^2)$ is the same for each block. A is then equal to $1/n$. Further, since each $(\sum_j u_{ij}^2)$ is essentially positive, it is seen that the maximum value A can have is when $(\sum_j u_{ij}^2)$ is zero in every block, except one. A is then equal to unity. Hence from (16) $\sigma_{U_R}^2$ must lie between $\dfrac{2(n-1)}{n^2(ns-n)}$ and 0. Comparing the maximum possible value, with the value $\dfrac{2(n-1)}{n^2(ns-n+2)}$ of $\sigma_{U_N}^2$ in equation (7), it appears that, if $n(s-1)$ is not too small, σ_{U_R} will never be much greater than σ_{U_N} and hence the chance of rejecting will never much exceed the specified ϵ. We are therefore not likely to err much on the side of overestimating the significance of observed treatment differences. On the other hand if there is too much discrepancy between the variances within the different blocks, σ_{U_R} may be considerably less than σ_{U_N} and the test may seriously underestimate significance. The question must now be asked: how much discrepancy in block variations will be serious?

Our procedure will be to approximate to the U_R-distribution by means of a Type I Pearson-Curve with limits at 0 and 1, i.e. by a curve

$$p(U_R) = \text{const.} \times U_R^{m_1-1}(1-U_R)^{m_2-1}. \qquad \ldots\ldots(17)$$

m_1 and m_2 will be chosen so that the first two moments of this curve agree with the true moments of U_R given by equations (12) and (14). Of course, the distribution of U from randomization must in fact be discontinuous and although U must lie between 0 and 1, these extreme values will never in general be attained. However, although (17) may for these reasons only provide an approximate graduation to the U_R distribution, we may certainly expect it to be better than the normal theory curve of (5), which is of the same form but does not have the correct standard deviation. Certainly for the normal theory to be satisfactory we may demand good correspondence between (5) and (17).

4. *A Particular Example* (*Randomized Blocks*). In the following example there are $n = 8$ blocks and $s = 4$ treatments. This is the case for which T. Eden and F. Yates performed the sampling experiment referred to earlier. f_0 is 21, f_1 is 3 and equation (5) gives

$$p(U_N) = \text{const.} \times U_N^{\frac{9}{2}-1}(1-U_N)^{\frac{9}{2}-1}.$$

From the *Tables of the Incomplete Beta-function** or by transformation from Fisher's z-tables, we find

$$P(U_N > \cdot 305) = \cdot 05; \quad P(U_N > \cdot 410) = \cdot 01.$$

That is to say, ·305 and ·410 are the values of U_0 corresponding to normal theory significance levels of $\epsilon = \cdot 05$ and $\epsilon = \cdot 01$ respectively. Now the m_1 and m_2 of the general type I curve, (17), are connected with its first two moments by the relations

$$m_1 = \frac{\mu_1'(\mu_1' - \mu_2')}{(\mu_2' - \mu_1'^2)}; \quad m_2 = \frac{(1-\mu_1')(\mu_1' - \mu_2')}{(\mu_2' - \mu_1'^2)}. \qquad \text{.....(18)}$$

Also in the present case the randomization moments from equations (12) and (14) are

$$\mu_1' = \tfrac{1}{8}; \quad \mu_2' = \frac{5 - 2A}{192}.$$

Hence

$$m_1 = \frac{(19+2A)}{16(1-A)}; \quad m_2 = \frac{7(19+2A)}{16(1-A)}.$$

For every A in the possible range from $\frac{1}{8}$ to 1 there will be a different approximation (17) to the distribution of U from randomization. For each A, then, we can derive, from the Incomplete Beta-function tables, an approximation to $P(U_R > U_0)$ —the true chance of the first kind of error. This has been done and the results are plotted in Fig. 2. It will first be noticed that the risk of the first kind of error never exceeds by much the value ϵ at which we attempt to fix it. The maximum value occurs at $A = \cdot 125$, where for $\epsilon = \cdot 05$ the risk is ·056 and for $\epsilon = \cdot 01$ it is ·013. The risk decreases as A increases, until at $A = \cdot 192$ it is actually ϵ. It decreases further from ϵ to 0 as A increases from ·192 to 1. For values of A within this range the test will tend to underestimate significance.

The data upon which T. Eden and F. Yates performed their sampling experiment were derived from measurements of heights of barley. The eight values of $\sum_j u_{ij}^2$ $(i = 1, 2, \ldots, s)$ for these data are 7628, 15,702, 22,669, 59,732, 3666, 90,593, 26,297 and 8672. By (15) these give $A = \cdot 242$. $\sigma_{U_R}^2$ is ·0079 against the normal theory value ·0084 of $\sigma_{U_N}^2$. From Fig. 2 the risks of the first kind of error are ·046 and ·0085 instead of ·05 and ·01. There is a slight tendency to underestimate significance, but for practical purposes this is negligible.

5. *Further Examples* (*Randomized Blocks*). Whether the test will usually be unbiased depends on the values of A which we are likely to meet in practice. An examination of uniformity trial data in different fields would therefore be of value. In the following I have considered four examples.

(I) A trial with mangolds by A. Mercer and W. Hall published in *Journ. Agric. Sci.* IV (1911), p. 107.

(II) A trial with wheat by A. Mercer and W. Hall in the same paper.

* *Tables of the Incomplete Beta-function*, edited by Karl Pearson, *Biometrika* Office, University College, London.

(III) A trial with oats on the field "Za Baranem" published in the Polish journal *Roczniki Nauk Rolniczych* (1917) and in *Landw. Versuchstationen*, XC (1917), pp. 225–40 (authors, M. Gorski and M. Stefaniow).

(IV) An experiment giving nitrogen content in barley carried out by St Barbacki and published in *Mémoires de l'Institut National Polonais d'Économie Rurale à Pulawy*, XIV, Nr. 213 (1933), pp. 106–57.

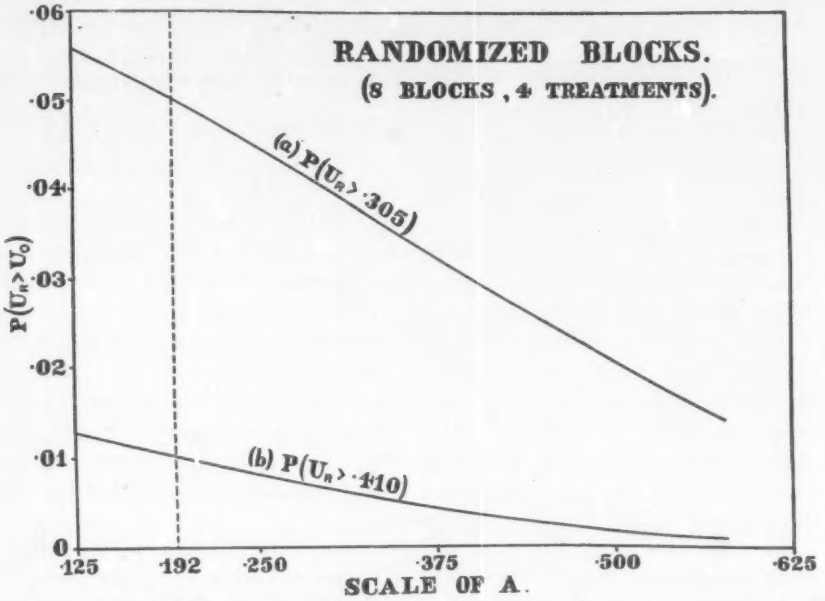


Fig. 2. True probability of the first kind of error in Randomized Block experiment with 8 blocks and 4 treatments. (*a*) Normal theory $\epsilon = \cdot 05$. (*b*) Normal theory $\epsilon = \cdot 01$.

In each case the data as published were grouped up until the plots were of such size and position as might be used in a Randomized Block experiment. For this amalgamated data the necessary information for comparing the U_N and U_R distributions is given in Table II.* It will be noted that for the first three examples there is exceedingly good agreement. For the fourth, the true risks of the first kind of error are ·041 and ·007 instead of ·05 and ·01, and the test tends to underestimate significance.

Whereas these practical trials show no serious bias in the test, it must not be inferred that this will always be the case. Theoretically it has been shown that the test will underestimate significance if the block variances are too discrepant.

* n and s are in each case different from 8 and 4, so that comparison should not be made with Fig. 2.

In the cases where the discrepancies are sufficiently large to matter, the experimenter will probably notice something peculiar in his data and make allowances accordingly. Sometimes, however, there may be doubt and an investigation on the above lines may be useful. It should be noted that the actual work involved is not great. The total "within block" variation is computed in any case for the analysis of variance, and it involves little extra trouble to calculate the separate "within block" variances, from which A is obtained. In this connection a table may be useful showing for different n and s the range of values, A, for which the bias in the test is negligible.

TABLE II

The comparison of U_N and U_R distributions for certain uniformity trial data

Example:	I	II	III	IV
$n \times s$	10×5	12×3	6×5	4×4
$\sigma^2_{U_N}$	·00426	·00588	·01068	·02679
A	·1312	·1847	·2023	·4258
$\sigma^2_{U_R}$	·00434	·00566	·01108	·02392
$p(U_R > U_0)$ for $\epsilon = ·05$	·050	·048	·053	·041
$p(U_R > U_0)$ for $\epsilon = ·01$	·010	·009	·011	·007

It is of interest to note that the procedure described above, when applied to the example of section 21 of R. A. Fisher's *Design of Experiments*, gives results consistent with his. This example relates to an actual experiment in which the question is asked whether one kind of seed is better than another. A positive difference of means was observed and from the t test there was a chance ·02491 of getting a result as great or greater without there really being a differentiation in seed. By considering the 2^{15} results, which could be obtained in this experiment by randomization on the "null" hypothesis, Fisher obtained, without any approximation, a significance level of ·02628 for the observed difference. By means of the Type I approximation to the U_R distribution I found that, corresponding to the normal theory $\epsilon = ·05$, the chance $p(U_R > U_0)$ for this data was ·053. As I was considering the chances of obtaining as large an absolute deviation as the one observed, the result agrees with Fisher's as far as the third decimal place.

6. *Latin Squares (Normal Theory).* In Latin Square experiments the field is divided into s rows ($i = 1, 2, \ldots, s$) and s columns ($j = 1, 2, \ldots, s$), making s^2 plots. The treatments tested ($k = 1, 2, \ldots, s$) are arranged on the field so that each falls once and once only into every row and every column. Upon the yields of the experiment the analysis of Table III is performed.

The test, whether there are significant differences between the treatment means, involves the calculation of $z = \frac{1}{2} \log_e (v_1/v_0)$ and reference to tables based on normal theory. This theory proceeds from the assumption that, if the treat-

ments are equivalent, the yield of the plot in the ith row and jth column can be written

$$x = A + R_i + C_j + \eta, \qquad \text{.....(19)}$$

where A, R_i $(i = 1, 2, \ldots, s)$ and C_j $(j = 1, 2, \ldots, s)$ are constants, and the η's are normally and independently distributed about zero with the same standard

TABLE III

Analysis of Variance for Latin Square

Source	Degrees of Freedom	Sum of Squares	Mean Square
Between Treatments	$(s-1)$	S_1	v_1
Between Rows	$(s-1)$	S_2	v_2
Between Columns	$(s-1)$	S_3	v_3
Residual	$(s-1)(s-2)$	S_0	v_0
Total	(s^2-1)		

deviation. As in the case of Randomized Blocks, we shall consider $U = (S_1)/(S_0 + S_1)$, which is now related to z by the equation

$$U = \{1 + (s-2)e^{-2z}\}^{-1}. \qquad \text{.....(20)}$$

On the assumption (19), we have

$$U_N = \frac{\chi_1^2 \sigma^2}{\chi_0^2 \sigma^2 + \chi_1^2 \sigma^2},$$

where χ_0^2 and χ_1^2 are independently distributed as χ^2 with degrees of freedom $f_0 = (s-1)(s-2)$ and $f_1 = (s-1)$ respectively. The distribution of U_N* is therefore

$$p(U_N) = \text{const.} \times U_N^{(f_1/2)-1}(1 - U_N)^{(f_0/2)-1}, \qquad \text{.....(21)}$$

and the moments are

$${}_N\mu_1' = \frac{1}{(s-1)}, \qquad \text{.....(22)}$$

$${}_N\mu_2' = \frac{(s+1)}{(s-1)(s^2-2s+3)}, \qquad \text{.....(23)}$$

$$\sigma_{U_N}^2 = \frac{2(s-2)}{(s-1)^2(s^2-2s+3)}. \qquad \text{.....(24)}$$

7. *Latin Squares* (*Randomization Theory*). It is now necessary to consider in more detail the arrangement of the experiment, and to formulate precisely the hypothesis which it is meant to test. Let the yield which the kth treatment is capable of giving on the plot (i, j) be $x_{ij(k)}$. Then we shall suppose that the hypothesis under test is that *every plot* would give the same yield, however treated, i.e. that

$$x_{ij(k)} = x_{ij} \qquad (k = 1, 2, \ldots, s). \qquad \text{.....(25)}$$

* The suffix N is used, as before, to denote the distribution of U on normal theory.

We are concerned only with the probability of the first kind of error, and therefore the whole of the following analysis supposes (25) satisfied. As has been stated, the experiment is arranged so that each treatment falls once into every row and every column. For example, suppose the yields x_{ij} form on the field the 4×4 square of Fig. 3 (a), and the arrangement of the treatments on the field follows the plan of Fig. 3 (b). Then the yields reclassified by treatment and row will be as in Fig. 3 (c). The treatment means are 20·25, 24·75, 23·75 and 23, and from them may be calculated the treatment sum of squares, S_1 of Table III. Any other arrangement of treatments, differing from Fig. 3 (b), but satisfying the Latin Square conditions, will lead to a different reclassification of the yields and a different S_1. Randomization enters into the experiment in making the decision as to which particular Latin Square arrangement is to be applied. A fundamental set of possible squares

		Columns (j)			
		1	2	3	4
Rows (i)	1	30	22	18	28
	2	34	13	21	25
	3	38	12	18	23
	4	33	19	16	17

Fig. 3 (a). Example of yields x_{ij} on a 4×4 field.

		Columns (j)			
		1	2	3	4
Rows (i)	1	4	2	1	3
	2	1	3	2	4
	3	3	1	4	2
	4	2	4	3	1

Fig. 3 (b). Example of Latin Square arrangement of treatments ($k = 1, 2, 3, 4$).

		Treatments (k)			
		1	2	3	4
Rows (i)	1	18	22	28	30
	2	34	21	13	25
	3	12	23	38	18
	4	17	33	16	19
Mean		20¼	24¾	23¾	23

Fig. 3 (c). Reclassification of yields obtained on applying Fig. 3 (b) to Fig. 3 (a).

(defined below) is decided upon, and from it one particular square is chosen at random for the experiment. In order to judge, therefore, the significance of the value of the criterion U obtained from an experiment, it is necessary to know something of the distribution of values of U which would be generated if every element of the fundamental set of Latin Squares were applied to the field under essentially the same conditions. As before, attention will here be confined to the first two moments of this distribution, and comparisons will be made with the normal theory moments of (22), (23) and (24).

First the fundamental set of squares must be defined. For s small, this set can be taken to consist of all the different Latin Squares that are possible. Methods of choosing one square at random from this total set have been given by R. A. Fisher and F. Yates for $s \leqslant 6$.* The necessary enumeration of squares, which would make these methods available for $s > 7$, has not yet been performed. Instead, for s

* See F. Yates, "The Formation of Latin Squares for Use in Field Experiments", *Emp. J. exp. Agric.* I (1933), pp. 235–44 and R. A. Fisher and F. Yates, "The 6×6 Latin Squares", *Proc. Camb. phil. Soc.* XXX (1934), pp. 492–507.

from 7 to 12, Yates has simply given a single example of a Latin Square, and suggests that the fundamental set of squares, to be used in randomization, should consist of all squares that can be obtained from this single square, by permutations of rows, columns and treatments. Such a set of squares has been called a *transformation* set.

The mean and second moments of U for squares of a given transformation set will first be considered. It will be found that, for a given s, the mean U for different transformation sets is the same, but this is not the case for the second moment. Knowing the second moment for each transformation set it is not difficult, in the cases when $s > 6$, to deduce the second moment over a fundamental set consisting of all squares that are possible. We shall first consider moments over single transformation sets.

It is convenient to write

$$\begin{aligned} x_{ij} &= x_{..} + (x_{i.} - x_{..}) + (x_{.j} - x_{..}) + (x_{ij} - x_{i.} - x_{.j} + x_{..})^* \\ &= A' + R'_i + C'_j + u_{ij} \quad \text{(say)}, \end{aligned} \qquad \ldots\ldots(26)$$

where it will be noted that

$$\left.\begin{aligned} &\sum_i u_{ij} = 0 \quad (j = 1, 2, \ldots, s); \qquad \sum_j u_{ij} = 0 \quad (i = 1, 2, \ldots, s); \\ &\sum_i R'_i = 0; \qquad\qquad \sum_j C'_j = 0. \end{aligned}\right\} \qquad \ldots\ldots(27)$$

Also if the yield of the kth treatment in the ith row is denoted by $x_{i(k)}$, we can write

$$x_{i(k)} = x_{..} + (x_{i.} - x_{..}) + (x_{.j} - x_{..}) + (x_{i(k)} - x_{i.} - x_{.j} + x_{..}),$$

where j is the column into which the kth treatment falls in the ith row. This means that we can write

$$x_{i(k)} = A' + R'_i + C'_j + y_{i(k)} \quad \text{(say)}. \qquad \ldots\ldots(28)$$

It will be seen that only variation in the quantities $y_{i(k)}$ need be considered in the following analysis. The possible sets of values $y_{i(k)}$, which can be obtained by applying the transformation set of squares to the field, are those which can be obtained by applying the squares to the residuals u_{ij}.

The numerator of U is

$$S_1 = s \sum_k \{x_{\cdot(k)} - x_{..}\}^2 = s \sum_k y^2_{\cdot(k)}. \qquad \ldots\ldots(29)$$

The denominator is

$$\begin{aligned} (S_0 + S_1) &= \sum_i \sum_j (x_{ij} - x_{..})^2 - \sum_i s\,(x_{i.} - x_{..})^2 - \sum_j s\,(x_{.j} - x_{..})^2 \\ &= \sum_i \sum_j (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 = \sum_i \sum_j u^2_{ij}, \end{aligned} \qquad \ldots\ldots(30)$$

i.e. is the same no matter what the Latin Square arrangement is. The moments of U depend therefore only on the moments of S_1, and these in turn depend only on the possible sets of y's. From (29) it is seen that S_1 is symmetrical with respect to

* Dots indicate that means are being taken over all the values of the letter replaced by the dot.

treatments. All Latin Squares of the transformation set, which can be derived from one another simply by a permutation of treatments, will therefore give rise to the same value of S_1 and therefore of U. Hence the moments of U over the transformation set are the same as the moments over a set, which can be derived from an original square by permutation of rows and columns only. We shall denote this set by Ω and expectations over Ω by E. Then from (29)

$$E(sS_1) = E\left[\sum_k \{\sum_i y_{i(k)}\}^2\right]$$

$$= E\left[\sum_k \sum_i y^2_{i(k)} + \sum_k \sum_{i \neq m} y_{i(k)} y_{m(k)}\right]$$

$$= \sum_i \sum_j u^2_{ij} + s \sum_{i \neq m} \left\{ \frac{\sum_{j \neq j'} u_{ij} u_{mj'}}{s(s-1)} \right\}, \quad *$$

since, for any treatment, the column j occupied in row i and the column j' in row m can be with equal likelihood any pair of values, except $j = j'$. Using the relations (27) we obtain

$$E(sS_1) = \sum_i \sum_j u^2_{ij} + s \sum_{i \neq m} \left\{ \frac{-\sum_j u_{ij} u_{mj}}{s(s-1)} \right\}$$

$$= \sum_i \sum_j u^2_{ij} + \frac{1}{(s-1)} \sum_j \sum_i u^2_{ij} = \frac{s(\sum_i \sum_j u^2_{ij})}{(s-1)}.$$

Hence
$$E(U) = E\left(\frac{S_1}{S_0 + S_1}\right) = \frac{1}{(s-1)}, \qquad \text{.....(31)}$$

agreeing with the normal theory value of (22). For the second moment we have

$$E(s^2 S_1^2) = E\left[\sum_k \{\sum_i y_{i(k)}\}^2\right]^2$$

$$= E\left[\sum_k \sum_{k'} \sum_i \sum_m \sum_p \sum_q y_{i(k)} y_{m(k)} y_{p(k')} y_{q(k')}\right]. \qquad \text{.....(32)}$$

The method used to evaluate this expectation in the case of Randomized Blocks can no longer be applied, since the y's in different rows are not independent. The difficulty is greatest for terms in which $k \neq k'$, e.g. the term $y_{i(1)} y_{m(1)} y_{p(2)} y_{q(2)}$. To obtain the expectation of this, it is convenient to divide the set Ω into a number of sub-sets and first find the expectation for each sub-set. We shall put into one sub-set all the squares of Ω in which treatment 1 is allocated to the same plots. Such a sub-set will be termed $\omega(j_1, j_2, \ldots, j_s)$, where j_i is the column into which treatment 1 falls in the ith row. For example, the square of Fig. 4 (a) is a member of $\omega(2, 1, 3, 5, 4)$, and all the members of this set will be obtained by permuting rows and columns in such a way that the 1's are not moved. This permutation may be done by interchanging rows in any way, and then making the necessary column permutation to bring the 1's back to their original positions. If, for

* For summation convention see footnote, p. 26.

example, rows 1 and 4 are interchanged, then necessarily columns 2 and 5 must be interchanged. For different sets ω, the column permutation to be taken in conjunction with any particular row permutation is, of course, different.

4	(1)	3	5	2
(1)	3	2	4	5
3	5	(1)	2	4
5	2	4	3	(1)
2	4	5	(1)	3

Fig. 4 (*a*). Example of 5×5 square belonging to ω (2, 1, 3, 5, 4).

Row M				
1	(1)	3	2	4	5
2	4	(1)	3	5	2
3	3	5	(1)	2	4
4	2	4	5	(1)	3
5	5	2	4	3	(1)

Fig. 4 (*b*). Square of same set Ω as Fig. 4 (*a*), but belonging to ω (1, 2, 3, 4, 5). This particular square may be chosen as the fundamental square of ω (1, 2, 3, 4, 5).

Instead of considering immediately expectations over the general sub-set $\omega(j_1, j_2, \ldots, j_s)$, it is simpler to start with the particular set $\omega(1, 2, \ldots, s)$, for which treatment 1 lies down the principal diagonal. Expectations over this set will be denoted by E'. Then

$$E'[\sum_i \sum_m \sum_p \sum_q y_{i(1)} y_{m(1)} y_{p(1)} y_{q(1)}] = \sum_i \sum_m \sum_p \sum_q u_{ii} u_{mm} u_{pp} u_{qq}$$
$$= (\sum_i u_{ii})^4, \qquad \ldots\ldots(33)$$

and
$$E'[\sum_i \sum_m \sum_p \sum_q y_{i(1)} y_{m(1)} y_{p(2)} y_{q(2)}] = (\sum_i \sum_m u_{ii} u_{mm})\, E'[\sum_p \sum_q y_{p(2)} y_{q(2)}]$$
$$= (\sum_i u_{ii})^2\, E'[\sum_p \sum_q y_{p(2)} y_{q(2)}]. \qquad \ldots\ldots(34)$$

Our first problem, therefore, is to evaluate terms of the form $E'[y_{p(2)} y_{q(2)}]$. For, having these, we can deduce (34); then by analogy we can obtain the expectations of (33) and (34) over the more general sub-set $\omega(j_1, j_2, \ldots, j_s)$; finally we can combine all the sub-sets* to obtain the expectations of the same quantities over Ω. $E(s^2 S_1^2)$ will then follow from (32).

To fix ideas we shall take one member of $\omega(1, 2, \ldots, s)$ and term it the fundamental square of the set. The rows in this square will be numbered $M = 1, 2, \ldots, s$. (Fig. 4 (*b*) shows this done for squares belonging to the same set Ω as the square of Fig. 4 (*a*).) The manner in which all the squares of the set can be obtained from the fundamental square is this: the rows can be permuted in any way: the columns must then be permuted in exactly the same way, in order that the 1's should come back on to the principal diagonal. For short, this type of permutation of rows and columns will be termed a symmetrical permutation.

* There will be the same number of different squares in each sub-set $\omega(j_1, j_2, \ldots, j_s)$.

First take $q=p$. Then making all symmetrical permutations, treatment 2 in row p will fall with equal frequency into all columns except the pth. Therefore

$$E'[y_{p(2)}^2]=(\sum_j u_{pj}^2-u_{pp}^2)/(s-1). \qquad \text{.....(35)}$$

Next consider $q\neq p$. Treatment 1 will occupy in the rows p and q the positions u_{pp} and u_{qq}. There are now (s^2-3s+3) possible pairs of values for $y_{p(2)}$ and $y_{q(2)}$, but all of these do not occur equally frequently in the set $\omega(1, 2, \ldots, s)$. The possible kinds of pairs are illustrated in the following diagram, the bracket denoting the plot on which treatment 2 falls.

Row p ... u_{pp} ... (u_{pq}) ... ... u_{pp} ... (u_{pq}) ... ...
Row q ... (u_{qp}) ... u_{qq} ... ... (u_{qj}) ... ... u_{qq} ... ...
(i) (ii)

Row p ... u_{pp} ... ... (u_{pj}) ... ... u_{pp} ... ... (u_{pj}) ...
Row q ... (u_{qp}) ... u_{qq} ... ... $(u_{qj'})$ ... ... u_{qq} ... ...
(iii) (iv)

We may have

(i) $y_{p(2)}=u_{pq}$ and $y_{q(2)}=u_{qp}$,
(ii) $y_{p(2)}=u_{pq}$ and $y_{q(2)}=u_{qj}$ $(j\neq p \text{ or } q)$,
(iii) $y_{p(2)}=u_{pj}$ and $y_{q(2)}=u_{qp}$ $(j\neq p \text{ or } q)$,
or (iv) $y_{p(2)}=u_{pj}$ and $y_{q(2)}=u_{qj'}$ $(j \text{ and } j'\neq p \text{ or } q,\ j\neq j')$.

There is 1 pair of kind (i), $(s-2)$ of kind (ii), $(s-2)$ of kind (iii), and $(s-2)(s-3)$ of kind (iv),* making (s^2-3s+3) possible values for $y_{p(2)}y_{q(2)}$. To find the relative frequency of these we must now consider certain properties of what we have termed the fundamental square of $\omega(1, 2, \ldots, s)$.

Take a particular row M of this square. In this row treatment 1 falls into column M. There will now be some row (R say) for which treatment 2 falls into column M.

	Column M		Column R	
Row M	1		?	
Row R	2		1	

* I am throughout taking s to be greater than 3.

In this row treatment 1 falls into column R. There are now two possibilities: either (a) treatment 2 in row M falls into column R, or (b) it does not. If it does, we shall say that treatments 1 and 2 satisfy, for row M, the reversal property P. If not, we shall say that they satisfy the property $\overline{P}$. Further, taking all the rows M ($=1, 2, \ldots, s$), we shall denote by n_{12} the number of them in which the property P is satisfied for treatments 1 and 2. More generally by $n_{kk'}$ will be denoted the number of rows in which treatments k and k' satisfy the reversal property. (For example, in the square of Fig. 4 (b), for the treatments 1 and 2, the property P holds for $M=2$ and 5, the property $\overline{P}$ for $M=1$, 3 and 4. Hence n_{12} is 2.)

The relevance of the above considerations is seen when we come to find the proportion of squares of ω $(1, 2, \ldots, s)$ which give $y_{p(2)}y_{q(2)}=u_{pq}u_{qp}$. This is the situation pictured in the first figure of the diagram above. For any square of the set ω $(1, 2, \ldots, s)$ which produces this situation, the four elements falling on the plots u_{pp}, u_{pq}, u_{qp} and u_{qq} satisfy, for treatments 1 and 2, the reversal property. These four elements must therefore correspond to four elements satisfying the reversal property somewhere in the fundamental square of ω $(1, 2, \ldots, s)$, for permutation of rows and columns will not destroy the property. The chance that $y_{p(2)}y_{q(2)}=u_{pq}u_{qp}$, therefore, depends on the number of times the property holds in the fundamental square. Now we have seen that the squares of ω $(1, 2, \ldots, s)$ are obtained by any permutation of rows followed by the same permutation of columns. There are $s(s-1)$ ways in which two rows of the fundamental square can be chosen to fall on the rows p and q of the field. It is seen that the number of these pairs of rows in which treatments 1 and 2 satisfy the reversal property P is n_{12}. Hence the chance of $y_{p(2)}y_{q(2)}=u_{pq}u_{qp}$ in the set ω $(1, 2, \ldots, s)$ is $n_{12}/s(s-1)$.

Let us denote by p_1, p_2, p_3 and p_4 the chances of getting in ω $(1, 2, \ldots, s)$ the individual pairs of values $y_{p(2)}$ and $y_{q(2)}$ referred to above as of kinds (i), (ii), (iii) and (iv). Then we have

$$\left.\begin{aligned} &p_1+(s-2)p_2=1/(s-1); \quad p_2=p_3; \\ &p_1+(s-2)p_2+(s-2)p_3+(s-2)(s-3)p_4=1. \end{aligned}\right\} \qquad \ldots\ldots(36)$$

The first of these relations follows from the fact that in ω $(1, 2, \ldots, s)$ the chance that $y_{p(2)}=u_{pq}$ is $1/(s-1)$. The second follows from considerations of symmetry and the third from the fact that the chances of all (s^2-3s+3) pairs must add up to unity. Using the value of p_1, which we have already evaluated, we obtain from (36)

$$\left.\begin{aligned} &p_1=n_{12}/s(s-1); \quad p_2=p_3=(s-n_{12})/s(s-1)(s-2); \\ &p_4=\{n_{12}+s(s-3)\}/s(s-1)(s-2)(s-3). \end{aligned}\right\} \qquad \ldots\ldots(37)$$

Now

$$\begin{aligned} E'[y_{p(2)}y_{q(2)}] &= p_1u_{pq}u_{qp}+p_2\sum_j{}' u_{pq}u_{qj} \\ &\quad +p_3\sum_j{}' u_{pj}u_{qp} \\ &\quad +p_4\sum_j{}' u_{pj}\sum_{j'}{}'' u_{qj'}, \end{aligned}$$

where $\sum_j'$ denotes summation of j over all values 1, 2, ..., s, excluding p and q, and $\sum_{j'}''$ denotes summation of j' over all values 1, 2, ..., s, excluding p, q and j. Performing these summations, remembering that $\sum_j u_{ij} = 0$, and substituting the values of p_1, p_2, p_3 and p_4 from (37), we obtain finally

$$E'[y_{p(2)} y_{q(2)}] = \frac{1}{(s-1)(s-2)}[u_{pq}u_{qq} + u_{pp}u_{qp} - u_{pq}u_{qp} + u_{pp}u_{qq} - \sum_j u_{pj}u_{qj}]$$
$$+ \frac{n_{12}}{s(s-1)(s-2)(s-3)}[(s-1)u_{pq}u_{qq} + (s-1)u_{pp}u_{qp}$$
$$+ (s^2-3s+1)u_{pq}u_{qp} + u_{pp}u_{qq} - \sum_j u_{pj}u_{qj}]. \qquad \text{.....(38)}$$

We have now obtained the expectations of $y_{p(2)}^2$ and $y_{p(2)}y_{q(2)}$ $(q \neq p)$ over the symmetrical permutation set which keeps treatment 1 down the major diagonal. Of the two, the latter depends, not only on the possible yields of the plots in rows p and q, but also on the structure of the squares in Ω. It does not matter which particular square of Ω is chosen to evaluate n_{12}, for the number of rows, in which treatments 1 and 2 satisfy the property P, is quite independent of any permutation of rows and columns (e.g. for any of the squares of Fig. 4. n_{12} is equal to 2).

Next, $E'[\sum_p \sum_q y_{p(2)} y_{q(2)}]$ will be considered, i.e. $\sum_p E'[y_{p(2)}^2] + \sum_{p \neq q} E'[y_{p(2)} y_{q(2)}]$. This will follow from the necessary summations of (35) and (38). These summations are simplified by the following relations:

$$\left.\begin{aligned}
&\sum_{p\neq q} u_{pp}u_{qq} = (\sum_p u_{pp})^2 - \sum_p u_{pp}^2; \\
&\sum_{p\neq q} u_{pp}u_{qp} = \sum_{p\neq q} u_{pq}u_{qq} = -\sum_p u_{pp}^2; \\
&\sum_{p\neq q} u_{pq}u_{qp} = \sum_p \sum_q u_{pq}u_{qp} - \sum_p u_{pp}^2; \\
&\sum_{p\neq q} \sum_j u_{pj}u_{qj} = -\sum_j \sum_p u_{pj}^2.
\end{aligned}\right\} \qquad \text{.....(39)}$$

Using these relations, we get finally

$$E'[\sum_p \sum_q y_{p(2)} y_{q(2)}] = \frac{1}{(s-1)(s-2)}[(\sum_p u_{pp})^2 - s\sum_p u_{pp}^2 - \sum_p \sum_q u_{pq}u_{qp}$$
$$+ (s-1)\sum_p \sum_j u_{pj}^2] + \frac{n_{12}}{s(s-1)(s-2)(s-3)}[(\sum_p u_{pp})^2 - s(s-1)(\sum_p u_{pp}^2)$$
$$+ (s^2-3s+1)(\sum_p \sum_q u_{pq}u_{qp}) + (\sum_p \sum_j u_{pj}^2)]. \qquad \text{.....(40)}$$

Substituting (40) into (34), we obtain

$$E'[\sum_i \sum_m \sum_p \sum_q y_{i(1)} y_{m(1)} y_{p(2)} y_{q(2)}] = \frac{1}{(s-1)(s-2)}[X - sY - Z + (s-1)(\sum_i \sum_j u_{ij}^2)W]$$
$$+ \frac{n_{12}}{s(s-1)(s-2)(s-3)}[X - s(s-1)Y + (s^2-3s+1)Z + (\sum_i \sum_j u_{ij}^2)W], \qquad \text{.....(41)}$$

where
$$\left.\begin{aligned} X&=(\sum_i u_{ii})^4, \quad Y=(\sum_i u_{ii})^2(\sum_i u_{ii}^2),\\ Z&=(\sum_i u_{ii})^2(\sum_i\sum_m u_{im}u_{mi}), \quad W=(\sum_i u_{ii})^2. \end{aligned}\right\} \qquad \text{.....(42)}$$

The expectations of (33) and (41) have been evaluated for the set having treatment 1 down the principal diagonal, i.e. the set for which treatment 1 in row i occupies the column i. For the more general set $\omega\,(j_1, j_2, \ldots, j_s)$, in which treatment 1 in row i falls on the plot (i, j_i) the expectations will be of exactly the same form, with X, Y, Z and W defined by the more general expressions:

$$\left.\begin{aligned} X&=(\sum_i u_{ij_i})^4, \quad Y=(\sum_i u_{ij_i})^2(\sum_i u_{ij_i}^2),\\ Z&=(\sum_i u_{ij_i})^2(\sum_i\sum_m u_{ij_m}u_{mj_i}), \quad W=(\sum_i u_{ij_i})^2. \end{aligned}\right\} \qquad \text{.....(43)}$$

Further, the same expectations for the total set Ω will be obtained by replacing X, Y, Z and W by $E[X]$, $E[Y]$, $E[Z]$ and $E[W]$—their expectations over all the set of permutations $(j_1, j_2, \ldots, j_s)$. Thus

$$E[\sum_i\sum_m\sum_p\sum_q y_{i(1)}y_{m(1)}y_{p(1)}y_{q(1)}]=E[X] \qquad \text{.....(44)}$$

and

$$\begin{aligned} E[\sum_i\sum_m\sum_p\sum_q y_{i(1)}y_{m(1)}y_{p(2)}y_{q(2)}]&=\frac{1}{(s-1)(s-2)}\{E[X]-sE[Y]\\ &-E[Z]+(s-1)(\sum_i\sum_j u_{ij}^2)\,E[W]\}+\frac{n_{12}}{s(s-1)(s-2)(s-3)}\{E[X]-s(s-1)E[Y]\\ &+(s^2-3s+1)E[Z]+(\sum_i\sum_j u_{ij}^2)\,E[W]\}. \qquad \text{.....(45)} \end{aligned}$$

Hence

$$\begin{aligned} E[s^2S_1^2]&=E[\sum_k\sum_{k'}\sum_i\sum_m\sum_p\sum_q y_{i(k)}y_{m(k)}y_{p(k')}y_{q(k')}]\\ &=\frac{s}{(s-2)}\{(s-1)E[X]-sE[Y]-E[Z]+(s-1)(\sum_i\sum_j u_{ij}^2)\,E[W]\}\\ &+\frac{(\sum_{k\neq k'} n_{kk'})}{s(s-1)(s-2)(s-3)}\{E[X]-s(s-1)E[Y]+(s^2-3s+1)E[Z]\\ &+(\sum_i\sum_j u_{ij}^2)\,E[W]\}. \qquad \text{.....(46)} \end{aligned}$$

The derivation of these expectations of X, Y, Z and W is given in an Appendix to this paper. The results only are presented here. They involve four symmetrical functions of the plot yields, viz.

$$\left.\begin{aligned} D&=\sum_i\sum_j u_{ij}^4, \quad F=(\sum_i\sum_j u_{ij}^2)^2,\\ G&=\{\sum_i(\sum_j u_{ij}^2)^2+\sum_j(\sum_i u_{ij}^2)^2\}, \quad H=\sum_i\sum_m(\sum_j u_{ij}u_{mj})^2. \end{aligned}\right\} \qquad \text{.....(47)}$$

With this notation,

$$\left.\begin{aligned} E[X] &= \frac{1}{s(s-1)(s-2)(s-3)}[s^2(s+1)D+3(s^2-3s+1)F \\ &\qquad -3s(s-1)G+6H]; \\ E[Y] &= \frac{1}{s(s-1)(s-2)}[s^2D+(s-1)F-sG]; \\ E[Z] &= \frac{1}{s(s-1)(s-2)(s-3)}[s^2(s+1)D+(s^2-3s+3)F \\ &\qquad -3s(s-1)G+2(s^2-3s+3)H]; \\ E[W] &= \frac{\sqrt{F}}{(s-1)}. \end{aligned}\right\} \quad \ldots\ldots(48)$$

Substituting (48) into (46),

$$E[s^2S_1^2] = \frac{1}{(s-1)(s-2)^2(s-3)}[2s^2(s-1)D+(s^4-4s^3+2s^2+6s-6)F$$
$$-2s(s^2-3s+3)G-2(s^2-6s+6)H]+\frac{(\sum_{k+k'} n_{kk'})}{\{s(s-1)(s-2)(s-3)\}^2}[2s^2(s-1)^2D$$
$$+2(2s^2-6s+3)F-2s(s-1)(s^2-3s+3)G+2(s^4-6s^3+13s^2-12s+6)H]. \quad \ldots\ldots(49)^*$$

The expectation of U^2 then follows from

$$E[U^2] = E\left[\frac{S_1^2}{(S_0+S_1)^2}\right] = \frac{E[S_1^2]}{(\sum_i \sum_j u_{ij}^2)^2} = \frac{E[s^2S_1^2]}{s^2F}. \quad \ldots\ldots(50)$$

In the case of Randomized Blocks the expectation of U^2 depended only on the size of the experiment and on a single function, A, of the plot yields. From algebraic considerations it was possible to show that there was an upper limit to the probability of the first kind of error when the z-test was applied. For the Latin Square the situation is much more complicated. $E[U^2]$ depends on three functions of the plot yields (viz. D/F, G/F and H/F), and also on the function, $(\sum_{k+k'} n_{kk'})$, of the structure of a typical square of Ω. No attempt is made here to make any definite statement, which will be independent of the values of these functions, about the probability of the first kind of error. It is possible, however, without much difficulty, to make a direct trial of the applicability of the theoretical z-distribution in any particular instance, by calculating out from the plot yields the quantities D, F, G and H. This has been done in the following section for the data of uniformity trials, and for some hypothetical data in which there are certain systematic fertility gradients.

8. *The* 4×4, 5×5 *and* 6×6 *Squares.* In applying (49) to particular cases, I shall make use of the methods of choosing squares summarized conveniently by F. Yates in the *Empire Journal of Experimental Agriculture*.† The squares

* For confirmation of this result see Appendix B. † *Loc. cit.*

derivable from a single $s \times s$ Latin square by permutation of rows, columns and treatments are called a transformation set. All the $(s!)^3$ permutations do not lead to different squares, but however many different squares there are, each will be repeated the same number of times. A square with 1, 2, 3, ..., s along the first row and 1, 2, 3, ..., s down the first column is called a reduced square. From a reduced square $(s!)(s-1)!$ different squares can be generated by permuting all the rows, except the first, and all the treatments. The number of different squares in a transformation set is equal to the number of different reduced squares in the set multiplied by $(s!)(s-1)!$ There is, in general, more than one transformation set for a given s, but the different transformation sets do not contain the same number of reduced squares. To make a random choice from all the different possible squares of size $s \times s$ (*giving each the same chance of being used*) it is necessary to give each transformation set a chance of being used proportional to the number of reduced squares it contains. In the following we shall consider the distribution of U, firstly over each of the transformation sets, and, secondly, over the whole set of different possible squares of size $s \times s$. Expectations over the set of all possible squares will clearly be weighted means of the expectations over the separate transformation sets—the weights being proportional to the numbers of reduced squares in the sets.

In the cases $s=4$, 5 and 6, equations (49) and (50) give

$$s=4, \quad \mu_2' = \frac{1}{192F}[50F+96D-56G+4H] + \frac{(\sum_{k \neq k'} n_{kk'})}{9216F}[22F+288D-168G+76H]; \qquad \text{.....(51)}$$

$$s=5, \quad \mu_2' = \frac{1}{1800F}[199F+200D-130G-2H] + \frac{(\sum_{k \neq k'} n_{kk'})}{360\,000F}[46F+800D-520G+292H]; \qquad \text{.....(52)}$$

$$s=6, \quad \mu_2' = \frac{1}{8640F}[534F+360D-252G-12H] + \frac{(\sum_{k \neq k'} n_{kk'})}{4\,665\,600F}[78F+1800D-1260G+804H]. \qquad \text{.....(53)}$$

When $s=4$ there are two transformation sets, an illustration of each being given in Fig. 5 (*a*). The first set contains 3 different reduced squares and the other only 1. For the first set $(\sum_{k \neq k'} n_{kk'}) = 16$. This is seen in the following way. Consider, say, treatments 2 and 4 in, say, the 3rd row. Treatment 2 falls into the 3rd column. Now see in what row treatment 4 falls into the 3rd column. It is the 2nd row. Now

see if in the 2nd row treatment 2 falls into the same column as treatment 4 does in the 3rd row. It does not. Therefore treatments 2 and 4 do not possess, for row 3, the reversal property, which I have referred to as P. Similarly, this property is seen not to hold for any row, and hence $n_{24}=n_{42}=0$. In the same fashion it is seen that all the other n's are 0, except $n_{12}=n_{21}=4$ and $n_{34}=n_{43}=4$. Thus $(\sum_{k\neq k'} n_{kk'})=16$. In the second set the reversal property holds for all pairs of treatments throughout. All the n's are 4 and thus $(\sum_{k\neq k'} n_{kk'})=48$. The second moment of U about zero for the two sets is therefore obtained by substituting into (51) the values $(\sum_{k\neq k'} n_{kk'})=16$ and 48, respectively. For the second moment over all possible squares we must take $\frac{3}{4}$ of the first result plus $\frac{1}{4}$ of the second, the ratios of the numbers of the reduced squares being as 3 : 1. This is the same thing as substituting into (51) the weighted mean of the $(\sum_{k\neq k'} n_{kk'})$'s, i.e.

$$(3\times 16+1\times 48)/4=24.$$

1	2	3	4
2	1	4	3
3	4	2	1
4	3	1	2

I

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

II

Fig. 5 (*a*). Illustrations of two 4×4 transformation sets.

1	2	3	4	5
2	1	4	5	3
3	5	1	2	4
4	3	5	1	2
5	4	2	3	1

I

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

II

Fig. 5 (*b*). Illustrations of two 5×5 transformation sets.

For the 5×5 squares there are again two transformation sets, illustrated in Fig. 5 (*b*). In the first set there are 50 reduced squares and $(\sum_{k\neq k'} n_{kk'})=16$. In the second set there are 6 reduced squares and $(\sum_{k\neq k'} n_{kk'})=0$. The weighted mean of $(\sum_{k\neq k'} n_{kk'})$ is $(50\times 16+6\times 0)/56$, i.e. $14\frac{2}{7}$. The second moment of U about zero, over the two sets and over all possible squares, is obtained by substituting respectively these three numbers for $(\sum_{k\neq k'} n_{kk'})$ in (52).

For the 6×6 squares there are 22 transformation sets. Yates illustrates only 17 of these, since the other 5 can be obtained by rotating 5 of the 17 through a right angle. For our purpose also it is unnecessary to distinguish between two sets of squares, one of which is the other rotated through a right angle. Such rotation does not affect $(\sum_{k\neq k'} n_{kk'})$. The numbers of reduced squares and $(\sum_{k\neq k'} n_{kk'})$ for each of the 17 sets is given in Table IV. Substitution into (53) gives the corresponding μ_2' for U. The least $(\sum_{k\neq k'} n_{kk'})$ is 0, the greatest 108, and the weighted mean $33\frac{9}{147}$.

TABLE IV

Summary of 6×6 *transformation sets*

Yates Index Number	I	II	III	IV	V	VI	VII	VIII	IX
Number of Reduced Squares	2160	1080	1080	1080	1080	540	540	720	360
$(\sum_{k \neq k'} n_{kk'})$	20	20	16	44	28	28	76	60	60
Yates Index Number	X	XI	XII	XIII	XIV	XV	XVI	XVII	
Number of Reduced Squares	180	240	120	60	40	72	36	20	
$(\sum_{k \neq k'} n_{kk'})$	36	36	36	36	0	60	60	108	

9. *Examples* (*Latin Squares*). In the following section the examples are numbered and described. It may here be noted that, although we may derive $s!(s-1)!$ different squares from a reduced square by complete permutation of treatments and permutation of all rows except the first, all these squares do not give a different U. For if the hypothesis that the treatments are equivalent is true, any permutation of treatments will not affect the between treatment sum of squares and therefore will not affect U. A reduced square therefore gives only $(s-1)!$ different U's. When $s=4$, this means that one transformation set gives $3 \times 3! = 18$ values and the other $1 \times 3! = 6$ values. The complete set of possible values is 24. For $s=4$, therefore, it is not difficult to work out all the possible values and derive therefrom second moments from randomization. This was done in the first example to test the correctness of (49).

Example I (4×4). Artificially constructed set of values u_{ij} of Fig. 6 (*a*).

*Example II** (4×4). Uniformity trial giving nitrogen content in barley (St Barbacki).

*Example III** (5×5). Uniformity trial with oats (Gorski and Stefaniow).

*Example IV** (6×6). Uniformity trial with wheat (Mercer and Hall).

Example V (6×6). An artificial set of yields x_{ij}, given in Fig. 6 (*b*), in which the fertility level runs diagonally across the field.

Example VI (6×6). An artificial set of yields x_{ij} given in Fig. 6 (*c*), in which the yield on any plot is equal to the yield on the plot two columns to the right in the next row.

For each of the above examples the necessary functions of the plot yields were computed and substituted in the appropriate equation (51), (52) or (53). For the 6×6 squares the equation (53) was not evaluated for all the transformation sets but only for the ones giving the extreme values of 0 and 108 to $(\sum_{k \neq k'} n_{kk'})$. The

* Examples II, III, IV are regroupings of the same data used in Examples IV, III and II of section 5 of the paper. The necessary references are given there. The Mercer and Hall wheat yields are given in Table VI.

expectation for the set of all possible squares was also evaluated. The results are given in Table V. The variance, $\sigma^2_{U_N}$, of U from normal theory is of course obtained from equation (24). In the last column but one is entered the ratio of the variance from randomization to the variance from normal theory, i.e. $\sigma^2_{U_R}/\sigma^2_{U_N}$.

1	0	3	−4
2	−2	0	0
3	1	−2	−2
−6	1	−1	6

(*a*)

50	41	32	17	65	71
65	50	41	32	17	65
47	65	50	41	32	17
23	47	65	50	41	32
38	23	47	65	50	41
41	38	23	47	65	50

(*b*)

50	41	32	17	65	77
65	59	50	41	32	17
46	51	65	59	50	41
24	19	46	51	65	59
35	29	24	19	46	51
32	47	35	29	24	19

(*c*)

Fig. 6. (*a*) Set of residuals u_{ij} of yields on artificial 4×4 field.
(*b*) and (*c*) Artificial sets of yields x_{ij} on 6×6 field.

TABLE V

Comparison of $\sigma^2_{U_R}$ and $\sigma^2_{U_N}$ for artificial examples and for data of three uniformity trials

Example	s	$\sum_{k+k'} n_{kk'}$	μ_2' from randomization	$\sigma^2{}_{U_R}$	$\sigma^2{}_{U_N}$	$\sigma^2{}_{U_R}/\sigma^2{}_{U_N}$	$P(U > U_0)$ for $\epsilon = \cdot 05$
I	4	16	·13914	·02803	·04040	·6938	—
		48	·13066	·01955	·04040	·4838	—
		24*	·13702	·02591	·04040	·6413	—
II	4	16	·13218	·02107	·04040	·5215	—
		48	·12417	·01306	·04040	·3233	—
		24*	·13018	·01907	·04040	·4720	—
III	5	16	·07809	·01559	·02083	·7481	—
		0	·07862	·01612	·02083	·7740	—
		$14\frac{2}{7}$*	·07814	·01564	·02083	·7509	·0288
IV	6	0	·048215	·008215	·011852	·6931	—
		108	·049106	·009106	·011852	·7683	—
		$33\frac{9}{147}$*	·048487	·008487	·011852	·7161	·0271
V	6	0	·05355	·01355	·011852	1·1432	—
		108	·05403	·01403	·011852	1·1839	—
		$33\frac{9}{147}$*	·05370	·01370	·011852	1·1556	·0624
VI	6	0	·05139	·01139	·011852	·9609	—
		108	·05415	·01415	·011852	1·1937	—
		$33\frac{9}{147}$*	·05223	·01223	·011852	1·0321	·0528

For each example two different transformation sets are considered and also the set of all possible squares. This latter is indicated by an asterisk in column 3. In the last column is given the probability of the first kind of error when the ϵ of normal theory is ·05.

For the uniformity trials II, III and IV it is seen that there is a very much greater disparity between the variances from the normal and randomization theories than was the case for Randomized Blocks. This seems to indicate that the z-test in the Latin Square is more liable to bias. Admittedly only three examples are considered here, but the data in two of them, at least, are not exceptional. Mercer and Hall's wheat data have often been used in discussions of the present character and I give in Table VI (a) the actual grouped yields which I used in Example IV. I made slight adjustments in these yields to make the sums of rows and columns divisible by 6. This made it simpler to evaluate the residuals $u_{ij}=(x_{ij}-x_{i.}-x_{.j}+x_{..})$. Some of the residuals turned out to be three-figure numbers, and as this makes the evaluation of D, F, G and H rather heavy I rounded them off to the two-figure numbers given in Table 6 (b).* These adjustments do not, I think, affect materially the ratios D/F, G/F and H/F.

TABLE VI

Mercer and Hall's wheat data

(a)

37·47	36·68	37·61	36·73	35·72	32·96
36·20	35·94	36·82	36·71	33·43	36·51
34·00	35·80	37·23	36·63	34·93	38·42
34·20	37·89	37·40	36·75	32·67	34·71
37·83	38·05	35·56	38·39	31·75	32·16
38·50	35·60	37·63	37·27	33·01	28·68

(b)

0·7	−0·4	0·2	−0·8	1·7	−1·4
−0·3	−0·9	−0·4	−0·5	−0·3	2·4
−2·8	−1·3	−0·2	−0·9	0·9	4·3
−2·0	1·4	0·5	−0·2	−0·8	1·1
1·6	1·5	−1·3	1·5	−1·7	−1·6
2·8	−0·3	1·2	0·9	0·2	−4·8

Mercer and Hall give wheat yields in lb. for 500 plots of $\frac{1}{500}$ acre each. By taking the first 18 rows and the first 18 columns of this data and regrouping in 36 bigger square plots of size $\frac{9}{500}$ acre, a 6×6 square with the yields in (a), above, was obtained. In (b) are given the residuals when row and column variation is allowed for. Certain adjustments made in deriving these residuals are referred to in the text.

It is an interesting point that the two systematic arrangements of Examples V and VI give good agreement between normal and randomization theories.

For the 5×5 and 6×6 squares here considered, the size of $(\sum_{k\neq k'} n_{kk'})$ does not seem to matter much. The expectations are much the same for all the transformation sets. For larger squares than $s=6$ the differences between transformation sets will probably become still less important.

In the last column of Table V I have given my approximation to the true probability of the first kind of error, when the rejection level is based on the 5 per cent. point U_0 of normal theory. These levels are obtained by approximating to the U distribution from randomization, by means of a Pearson type I curve (as in the case of Randomized Blocks). I have not done this in the examples with $s=4$, since there are only 24 possible values of U and the approximation by a continuous curve does not seem justified. In the other examples I give the risk of rejecting only when the randomization set consists of all possible squares of the size.

* I did the same thing with Examples II and III. Without these simplifications the numbers would have become uncomfortably large.

Results very little different would be obtained for the individual transformation sets. For the uniformity trials III and IV the probabilities of the first kind of error are ·029 and ·027 respectively, instead of the required ·05. There is in these cases a definite underestimation of significance by the usual z-test. This statement must, however, be qualified by the remarks in the next section.

10. *Summary and Conclusions.* In experiments in which randomization is performed, the actual arrangement of treatments on the field is one chosen at random from a predetermined set of possible arrangements. In the present paper investigation has been made for Randomized Blocks and Latin Square experiments, into the distribution of the statistic z, generated by the application to the observed plot yields of the whole fundamental set of arrangements, assuming as true the "null" hypothesis that the treatments have no differential effect on the plots. It was found convenient to consider, instead of z, a monotonically increasing function U of z, which is equal to the treatment sum of squares divided by the total of the treatment sum of squares and the residual sum of squares.

Comparison of the U distribution from randomization with that from normal theory showed, in both Randomized Blocks and Latin Square, exact agreement of the means, but disagreement in the variances with consequent disagreement in the proportions of the distribution falling beyond certain points. Some uniformity trial data were used in order to see whether, in practice, these disagreements were of sufficient magnitude to be of importance. For Randomized Blocks the cases considered showed close enough agreement between the randomization and normal theory variances of U. In each of three uniformity trials for Latin Square, however, the randomization variance of U was considerably smaller than that of the normal theory. Whether this should be taken as evidence of bias in the usual z-test based on normal theory, depends on the point of view adopted concerning the hypothetical population about which the data of an experiment is supposed to give information. Let us consider two possibilities.

(i) We may make, from the yields of the experiment, a *statistical* inference only about the situation on the particular field of the experiment, e.g. as in the present paper, we may be using our *statistical* method only to test whether all the treatments would have given identical yields on each plot of this particular field. Of course if we come to the conclusion that the treatments can be regarded as equivalent on this field, we probably make the further induction that they can be regarded as equivalent over some wider range of experience. Otherwise the experiment would be useless. However, from the present viewpoint, this further inference is not a statistical one in the usual sense. If the statistical part of the inferences made from an experiment go no farther than the experimental field, then in cases where the variance of U from randomization on the "null" hypothesis is less than that from normal theory, we may say that the usual z-test underestimates

the significance of the observed treatment differences. Where the variance of U from randomization is the larger, there the z-test overestimates significance.

(ii) We may, alternatively, choose to regard the inferences, drawn from the experimental yields to some wider experience, as being *completely* statistical. This means that we shall not only regard the statistic z calculated from the yields as a random sample from the distribution, which is obtained on the "null" hypothesis by applying all the possible arrangements of the fundamental set to the experimental field; we shall in turn regard this randomization distribution of z as a random sample from a set of similar distributions, which might be obtained in other experiments of a similar type. These may be carried out on different fields, under differing weather conditions, and may be subject to different technical errors in the harvesting and weighing of the crop and so on. Clearly with this wider conception, the results derived here for three Latin Square uniformity trials are insufficient to give any general answer to the question of bias arising in the application of the z-test. We should note, however, that randomization ensures the agreement of the mean U with normal theory. The second moment of U for *individual* fields may, as we have seen, differ appreciably from the normal theory value. In a number of experiments, however, these differences may tend to balance out, so that on the average the discrepancy may be negligible and the normal theory test unbiased.

In this connection it is of interest to recall the investigation of O. Tedin,* who considered the application of 5×5 Latin Squares to 91 uniformity trials. He took twelve different arrangements of the 5×5 square. Each arrangement was applied to all the 91 trials, giving for each arrangement 91 values of a criterion, which is practically the same as my U and which he termed a "treatment error coefficient". He came to the conclusion that it was dangerous to apply systematically the same arrangement (at least if it was of either the Diagonal or the Knight's Move pattern) in every experiment and still expect the normal theory z-test to be unbiased. The application of the methods of the present paper to such a set of uniformity trials would, I think, be useful. It would indicate how far the process of randomization does actually eliminate bias, when the z-test is regarded from the second viewpoint mentioned above.

APPENDIX

A. *Derivation of Expectations in Equation* (48). In the following, $\sum'$ will be used to denote summation over all possible sets of values of the row suffixes, excluding terms in which two or more of the row suffixes are the same. Thus, for example, $\sum' u_{ij_i}^2 u_{lj_l} u_{mj_m}$ is a summation over all i, l and m, excluding terms in

* O. Tedin, "The Influence of Systematic Plot Arrangements upon the Estimate of Error in Field Experiments", *J. agric. Sci.* XXI (1931), p. 191.

which two or more of i, l and m are equal. This summation, therefore, has $s(s-1)(s-2)$ terms. With this notation,

$$W=(\sum_i u_{ij_i})^2=\sum' u_{ij_i}^2+\sum' u_{ij_i}u_{lj_l};$$

$$Y=(\sum_i u_{ij_i})^2(\sum_i u_{ij_i}^2)=\sum' u_{ij_i}^4+\sum' u_{ij_i}^2u_{mj_m}^2+2\sum' u_{ij_i}^3u_{lj_l}+\sum' u_{ij_i}u_{lj_l}u_{mj_m}^2;$$

$$X=(\sum_i u_{ij_i})^4=\sum' u_{ij_i}^4+4\sum' u_{ij_i}^3u_{lj_l}+3\sum' u_{ij_i}^2u_{lj_l}^2+6\sum' u_{ij_i}^2u_{lj_l}u_{mj_m}+\sum' u_{ij_i}u_{lj_l}u_{mj_m}u_{rj_r};$$

$$\begin{aligned}Z=(\sum_i u_{ij_i})^2(\sum_i\sum_l u_{ij_l}u_{lj_i})&=\sum' u_{ij_i}^4+2\sum' u_{ij_i}^3u_{lj_l}+\sum' u_{ij_i}^2u_{mj_m}^2\\&\quad+2\sum' u_{ij_i}^2u_{ij_r}u_{rj_i}+\sum' u_{ij_i}^2u_{mj_r}u_{rj_m}+\sum' u_{ij_i}^2u_{lj_l}u_{mj_m}\\&\quad+2\sum' u_{ij_i}u_{lj_l}u_{ij_l}u_{lj_i}+4\sum' u_{ij_i}u_{lj_l}u_{ij_r}u_{rj_i}\\&\quad+\sum' u_{ij_i}u_{lj_l}u_{mj_r}u_{rj_m}.\end{aligned}\qquad\text{.....(54)}$$

W, X, Y and Z are thus dependent on twelve kinds of term, which are listed in column 1 of Table VII. The expectations of these terms are all derived in the same manner, and, as the algebra is somewhat long, only one example will be given in full here. For instance, consider $E(u_{ij_i}^2u_{mj_r}u_{rj_m})$. The expectation is being taken over all sets of values $(j_1, j_2, \ldots, j_s)$ which are permutations of the numbers $(1, 2, \ldots, s)$. The term under present consideration involves only the three different rows i, m and r, and we have only to consider what happens when j_i, j_m and j_r take all the values $(1, 2, \ldots, s)$, excluding any two of them being equal. Let us start by taking j_i and j_r fixed. Then j_m can take all values $(1, 2, \ldots, s)$, except j_i and j_r. Hence

$$\begin{aligned}E[u_{ij_i}^2u_{mj_r}u_{rj_m}]&=E\left[u_{ij_i}^2u_{mj_r}\frac{(\sum_j u_{rj}-u_{rj_i}-u_{rj_r})}{(s-2)}\right]\\&=-\frac{1}{(s-2)}\{E[u_{ij_i}^2u_{rj_i}u_{mj_r}]+E[u_{ij_i}^2u_{mj_r}u_{rj_r}]\}.\end{aligned}$$

Now consider j_i fixed, so that j_r can take all values $(1, 2, \ldots, s)$ except j_i. Hence

$$\begin{aligned}E[u_{ij_i}^2u_{mj_r}u_{rj_m}]=-\frac{1}{s-2}\Bigg\{&E\left[u_{ij_i}^2u_{rj_i}\frac{(\sum_j u_{mj}-u_{mj_i})}{(s-1)}\right]\\&+E\left[u_{ij_i}^2\frac{(\sum_j u_{mj}u_{rj}-u_{mj_i}u_{rj_i})}{(s-1)}\right]\Bigg\},\end{aligned}$$

i.e. $$E[u_{ij_i}^2u_{mj_r}u_{rj_m}]=\frac{1}{(s-1)(s-2)}\{E[2u_{ij_i}^2u_{rj_i}u_{mj_i}]-(\sum_j u_{mj}u_{rj})E[u_{ij_i}^2]\}.$$

Finally, j_i can take all values $(1, 2, \ldots, s)$. Hence

$$E[u_{ij_i}^2u_{mj_r}u_{rj_m}]=\frac{1}{s(s-1)(s-2)}\{2\sum_j u_{ij}^2u_{rj}u_{mj}-(\sum_j u_{inj}u_{rj})(\sum_j u_{ij}^2)\}.$$

The same method leads to the other entries in column 2, the essential point of the work being the repeated use of the relation $\sum_j u_{ij}=0$. Next we have to consider

TABLE VII

(1)	(2) Expectation of (1)	(3) Expectation of (1) summed over Σ'
$u^2_{ij_i}$	$\frac{1}{s}(\sum_j u^2_{ij})$	$\frac{1}{s}\sqrt{F}$
$u_{ij_i}u_{ij_l}$	$-\frac{1}{s(s-1)}(\sum_j u_{ij}u_{lj})$	$\frac{1}{s(s-1)}\sqrt{F}$
$u^4_{ij_i}$	$\frac{1}{s}(\sum_j u^4_{ij})$	$\frac{1}{s}D$
$u^3_{ij_i}u_{ij_l}$	$-\frac{1}{s(s-1)}(\sum_j u^3_{ij}u_{lj})$	$\frac{1}{s(s-1)}D$
$u^2_{ij_l}u^2_{ij_i}$	$\frac{1}{s(s-1)}\{(\sum_j u^2_{ij})(\sum_j u^2_{lj})-(\sum_j u^2_{ij}u^2_{lj})\}$	$\frac{1}{s(s-1)}\{F+D-G\}$
$u^2_{ij_i}u_{lj_l}u_{mj_m}$	$\frac{1}{s(s-1)(s-2)}\{2\sum_j u^2_{ij}u_{mj}u_{lj}-(\sum_j u_{lj}u_{mj})(\sum_j u^2_{ij})\}$	$\frac{1}{s(s-1)(s-2)}\{4D+F-2G\}$
$u^2_{ij_i}u_{ij_r}u_{rj_i}$	$-\frac{1}{s(s-1)}(\sum_j u^3_{ij}u_{rj})$	$\frac{1}{s(s-1)}D$
$u^2_{ij_i}u_{mj_r}u_{rj_m}$	$\frac{1}{s(s-1)(s-2)}\{2\sum_j u^2_{ij}u_{rj}u_{mj}-(\sum_j u_{mj}u_{rj})(\sum_j u^2_{ij})\}$	$\frac{1}{s(s-1)(s-2)}\{4D+F-2G\}$
$u_{ij_i}u_{lj_l}u_{ij_l}u_{lj_i}$	$\frac{1}{s(s-1)}\{(\sum_j u_{lj}u_{ij})^2-(\sum_j u^2_{ij}u^2_{lj})\}$	$\frac{1}{s(s-1)}\{H+D-G\}$
$u_{ij_i}u_{lj_l}u_{ij_r}u_{rj_i}$	$\frac{1}{s(s-1)(s-2)}\{2\sum_j u^2_{ij}u_{rj}u_{lj}-(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj})\}$	$\frac{1}{s(s-1)(s-2)}\{4D+H-2G\}$
$u_{ij_i}u_{lj_l}u_{mj_m}u_{rj_r}$	$\frac{1}{s(s-1)(s-2)(s-3)}\{(\sum_j u_{lj}u_{mj})(\sum_j u_{ij}u_{rj})+(\sum_j u_{ij}u_{lj})(\sum_j u_{mj}u_{rj})+(\sum_j u_{ij}u_{mj})(\sum_j u_{lj}u_{rj})-6\sum_j u_{ij}u_{lj}u_{mj}u_{rj}\}$	$\frac{3}{s(s-1)(s-2)(s-3)}\{2H+F+12D-6G\}$
$u_{ij_i}u_{lj_l}u_{mj_r}u_{rj_m}$	$\frac{1}{s(s-1)(s-2)(s-3)}\{(\sum_j u_{lj}u_{mj})(\sum_j u_{ij}u_{rj})+(\sum_j u_{ij}u_{lj})(\sum_j u_{mj}u_{rj})+(\sum_j u_{ij}u_{mj})(\sum_j u_{lj}u_{rj})-6\sum_j u_{ij}u_{lj}u_{mj}u_{rj}\}$	$\frac{3}{s(s-1)(s-2)(s-3)}\{2H+F+12D-6G\}$

the expectations of the elements in column 1, when the summation $\sum'$ is applied to them. This involves the evaluation of summations $\sum'$ of the quantities listed in column 1 of Table VIII. Again, only one example will be given here in full, to

TABLE VIII

(1)	(2) Column (1) summed over Σ'
$(\sum_j u^2_{ij})$	$\sqrt{F}$
$(\sum_j u_{ij} u_{lj})$	$-\sqrt{F}$
$(\sum_j u^4_{ij})$	D
$(\sum_j u^3_{ij} u_{lj})$	$-D$
$(\sum_j u^2_{ij})(\sum_j u^2_{lj})$	$F-G'$
$(\sum_j u^2_{ij} u^2_{lj})$	$G''-D$
$(\sum_j u^2_{ij} u_{mj} u_{lj})$	$2D-G''$
$(\sum_j u^2_{ij})(\sum_j u_{lj} u_{mj})$	$2G'-F$
$(\sum_j u_{lj} u_{ij})^2$	$H-G'$
$(\sum_j u_{ij} u_{lj})(\sum_j u_{ij} u_{rj})$	$2G'-H$
$(\sum_j u_{ij} u_{lj})(\sum_j u_{mj} u_{rj})$	$2H+F-6G'$
$(\sum_j u_{ij} u_{lj} u_{mj} u_{rj})$	$3G''-6D$

illustrate the method employed. The essential feature is the repeated use of the relation $\sum_i u_{ij}=0$. For instance, take $\sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj})$. First keep i and l fixed and sum r over all values $(1, 2, \ldots, s)$, excluding i and l:

$$\sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj}) = \sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}\overline{-u_{ij}-u_{lj}})$$
$$= -\sum'(\sum_j u_{ij}u_{lj})(\sum_j u^2_{ij}) - \sum'(\sum_j u_{ij}u_{lj})^2.$$

Now keep i fixed and sum l over all values $(1, 2, \ldots, s)$, excluding i:

$$\sum'(\sum_j u_{ij}u_{lj})(\sum_j u_{ij}u_{rj}) = -\sum'(\sum_j u^2_{ij})(\sum_j u_{ij}\overline{-u_{ij}}) - \{\sum_i\sum_l(\sum_j u_{ij}u_{lj})^2 - \sum_i(\sum_j u^2_{ij})^2\}$$
$$= 2\sum_i(\sum_j u^2_{ij})^2 - \sum_i\sum_l(\sum_j u_{ij}u_{lj})^2.$$

The results of the other similar summations are given in column 2 of Table VIII, use being made of the following notation:

$$D=\sum_i\sum_j u^4_{ij}, \qquad F=(\sum_i\sum_j u^2_{ij})^2,$$
$$G'=\sum_i(\sum_j u^2_{ij})^2, \qquad G''=\sum_j(\sum_i u^2_{ij})^2, \qquad H=\sum_i\sum_m(\sum_j u_{ij}u_{mj})^2,$$
$$G=G'+G''=\{\sum_i(\sum_j u^2_{ij})^2+\sum_j(\sum_i u^2_{ij})^2\}.$$

From the expressions of columns 2 of Tables VII and VIII, the expectations given in column 3 of Table VII are deduced. These are the expectations of the different kinds of terms involved in W, X, Y and Z, and by substitution into (54) the expectations $E[W]$, $E[X]$, $E[Y]$ and $E[Z]$ are obtained. These are given in equations (48) of the paper.

B. *Confirmation of Equation* (49). The algebraic processes leading to equation (49) are so heavy that one would feel more confident of their correctness if some practical test were made. This can very easily be done when $s=4$. For then there are only 24 possible values of S_1 and these can be calculated directly from the data. This was done in Example I of section 9 and exact agreement with the theory was observed. For $s>4$ the number of possible Latin Squares seemed too large to permit a complete investigation of this kind. It is possible, however, to obtain some general confirmation in the following way.

In the Randomization theory we made no assumptions about x_{ij}. Let us now consider what happens if we apply the reasoning of the theory to a situation where the x's do actually satisfy the equation

$$x_{ij}=A+R_i+C_j+\eta_{ij},$$

the η's being normal independent variates with mean zero and common standard deviation σ. One set of values x_{ij} satisfying these conditions may be termed a *configuration*. There are possible an infinite number of such configurations. We shall denote expectations over all these by E''.

Now consider the set Ω of possible Latin Square arrangements which can be applied to the values x_{ij}. Whatever individual square of Ω is applied to the set x_{ij} it is clear that, in repeated configurations, the resultant values of S_1 will be distributed as $\chi^2\sigma^2$ and therefore $E''[S_1^2]=(s^2-1)\sigma^4$. Hence, if $E[S_1^2]$ denotes the expectation of S_1^2 over all the Latin Squares of Ω applied to the *same* configuration, we must have *a fortiori*

$$E''\{E[S_1^2]\}=(s^2-1)\sigma^4, \qquad \text{i.e.} \quad E''\{E[s^2S_1^2]\}=s^2(s^2-1)\sigma^4.$$

But $E[s^2S_1^2]$ is given by the right-hand side of (49). Hence, if we take E'' {right-hand side of (49)}, we should obtain $s^3(s^2-1)\sigma^4$. Now it can be shown that

$$E''\{D\}=E''\{\sum_i\sum_j(x_{ij}-x_{i.}-x_{.j}+x_{..})^4\}=3(s-1)^4\sigma^4/s^2,$$

$$E''\{F\}=(s-1)^2(s^2-2s+3)\sigma^4,$$

$$E''\{G\}=2(s-1)^3(s+1)\sigma^4/s, \quad E''\{H\}=(s-1)^2(2s-1)\sigma^4.$$

Substituting these for D, F, G, H in (49), we do in fact get $s^2(s^2-1)\sigma^4$. This provides a check on the accuracy of the formula, although, of course, it does not constitute a proof of its correctness.

SOME ASPECTS OF THE PROBLEM OF RANDOMIZATION

By E. S. PEARSON

1. Introductory

The practical problem of mathematical statistics is to provide a conceptual model which will be of value to the man who needs to draw conclusions from the data of observation. In handling statistical data one of the commonest problems to be faced is that of drawing inferences from a part to the whole, from a sample to the population; such inferences are uncertain inferences, and it follows not only that in such cases the conceptual model must be constructed with the aid of the theory of probability but that its value to the practical man will be to some extent psychological. An historical study of the development of mathematical statistics shows an ever-increasing complexity in the structure of the abstract model and also an evolution of ideas as to how that model is to be of most use in practical application. In this course of evolution it is inevitable that many different suggestions should have been thrown out by mathematical statisticians as to the best way of linking the world of concepts with the world of experience. Ultimately, it is likely that the practical scientist, who may know relatively little mathematics but has to apply the methods of statistics in his research work, will play the decisive part in determining the form in which the theory of probability may be applied most usefully in different situations as a guide to judgment. But in the meantime it is necessary that amid the growing complication of the mathematical background statisticians should attempt to keep clear the simple principles which in their view have the greatest claim for acceptance.

An example of the gradual evolution of ideas is found in the changing attitude with which tests of goodness of fit and tests to determine whether differences are "significant" have been regarded. Perhaps one may say that 20 or 30 years ago the question posed by the statistician in applying such tests was often somewhat as follows:

"If my sample had come (*a*) from the population represented by my fitted curve, or (*b*) from a population whose parameters had the values given by the sample (and these estimates obtained from the sample cannot be very different from the unknown population values), what is the probability that a difference as great or greater than that observed would have occurred?"

It will be seen that the situation posed was to some extent hypothetical, since in fact the population sampled was not represented by the sample values. Nevertheless, the probability measure, P, obtained as an answer to this question

seemed to give the measure of assurance needed to make a decision. In so far as each problem was considered in isolation from other similar problems, the basis for any decision taken was to a large extent psychological.

In recent years we can follow the gradual introduction of a somewhat different conception. Its origin may be traced partly to the application of statistical methods in new fields where decisions had to be taken on evidence supplied by small samples, so that the differences between population values and sample estimates became so large that the hypothetical situation referred to above was seen to be noticeably unreal; and partly to the fact that in agricultural research investigations precisely similar tests were being applied again and again to the same type of experiment. Thus the relation was emphasized between (*a*) the probability measure, P, leading to a decision in an individual experiment, and (*b*) the expected proportion of times that a hypothesis of "no difference" would be wrongly rejected in the routine work of a research station. In terms of the older approach there might be little difference in an isolated problem between the psychological reaction to a P of ·05 and a P of ·02. But where experimental procedures were being repeated continually, the difference between a risk of mistake of 1 in 20 and of 1 in 50 might be of some consequence.*

Emphasis was therefore given in statistical literature to a new idea; that of planning a sampling procedure and the subsequent analysis of the data collected, in such a way as to control at any desired level the risk of making a wrong decision—that risk which can never be entirely eliminated in any form of work involving sampling. This change in attitude is illustrated by the form which many recently constructed probability tables have taken, following R. A. Fisher's suggestion. Instead of providing the statistician with the precise value of a probability measure, P, which he needed when regarding each problem in isolation, these tables are arranged so as to enable him to discover whether his test criterion falls below a certain "probability level", e.g. a 10, 5, 2, or 1 per cent. level. If then, for example, as a usual practice he rejects the hypothesis he is testing when the criterion falls below the 1 per cent. level (but not otherwise), he knows that in the long run of his experience this action will lead to one wrong decision in every hundred, a frequency of error which he may be quite prepared to accept.

This form of introduction of abstract theory into the world of experience has an obvious appeal to the practical man. If you tell him that theory enables him to assess the probability of a certain event in an individual trial or even to assess the frequency with which it would occur under somewhat hypothetical conditions, he may be unconvinced of the value of this theory to him. But if you can illustrate the statistician's objective by two examples of the following type, you are much more likely to convince him of the value of statistical tools.

* This has been brought out very clearly in questions of routine sampling in industry.

Example 1. A frequent problem is one in which, having a sample of n values of a variable x, it is wished to determine limits between which the unknown population mean, ξ, almost certainly lies. Under certain conditions the statistician can here provide a rule for determining from the sample data two limits ξ_a and ξ_b, such that the statement

$$\xi_a \leqslant \xi \leqslant \xi_b$$

may be made with a specified measure of confidence. The details of the procedure advocated depend on the application of the twofold principle that in making such a statement we are concerned, (*a*) to know the percentage of times it will be correct in long-run application under appropriate conditions, (*b*) to make the interval in some way as narrow as possible. To reduce the risk of error and to reduce the breadth of interval are, beyond a certain point, conflicting objectives and a balance must be struck between them; the statistical method shows how this may be done.

Example 2. Another common problem is one in which two samples are available and it is wished to test the hypothesis that they have been drawn from populations having the same means, $\xi_1 = \xi_2$. Again, the statistician can under certain conditions give a rule of procedure suggesting when the hypothesis should be rejected, and he may base this on another twofold principle: arrange so that in the long run application of the rule, (*a*) the hypothesis of "no difference" in means will only be rejected when it is true on a small and known percentage of occasions; (*b*) the hypothesis will be rejected as often as possible when there is a true difference in means, i.e. when $\xi_1 - \xi_2 \neq 0$.

These conceptions have no doubt always been present in the minds of mathematical statisticians but they have only been given precise formulation in recent years. The principle illustrated in Example 1 forms the basis of J. Neyman's work on confidence intervals and the confidence coefficient (1), and although presented in somewhat different form, I think, underlies R. A. Fisher's conception of fiducial probability (2). The principle mentioned in Example 2 forms the basis of J. Neyman and the present writer's work on the testing of statistical hypothesis (3), (4), but in the application of the conception (*b*) we are at variance with R. A. Fisher. In our view, just as in the simple problem of "interval estimation" mentioned in Example 1, it is necessary to specify the form of population distribution before the interval ξ_a, ξ_b can be calculated, so it is only possible to determine in any precise manner which is the most efficient test of a hypothesis if we can specify the class of alternative hypotheses. Thus, following quite simple principles, we may construct in the conceptual workshop the tests most appropriate in different precisely defined situations. It is then for the practical man to decide which of these situations corresponds most closely to that with which he is faced.

In Fisher's view the experimenter cannot and need not define the alternatives to the hypothesis he is testing. Indeed, Fisher would seem to consider it to be

important to a test of significance that it should be free from the necessity of introducing any elaborate type of background or alternatives which might be true. While I agree that the experimenter cannot specify all conceivable alternatives to the hypothesis tested, I think that a study of the situations met with in practice suggests that he does in fact usually have a fairly clear idea of the alternatives most likely to be true, and that if the mathematical statistician enables him to use this knowledge in picking out the most efficient statistical tool, he will be grateful. If it can be shown that in the situation most likely to exist (e.g. normal variation) one test will detect the falsity of the hypothesis of "no difference" more often than any other test, the appeal in favour of its adoption will surely be very strong.

2. Randomization

I have referred to the idea of arranging a sampling procedure so that conclusions drawn upon application of an appropriate statistical technique will be subject to a known and controlled risk of error. The principle of randomization, whose introduction is largely due to R. A. Fisher, provides a device to aid in the achievement of this objective. Most of the statistical tests used in the more complex sampling problems have been developed on the assumption that the variables are normally distributed, and while it is often clear that considerable departure from normality will not seriously effect their validity, it may be asked how far can tests be constructed which are completely independent of any assumption of normality?

Fisher has given an interesting illustration of such a test based on randomization in section 21 of his book, *The Design of Experiments* (5). The example is suggested by an investigation of Darwin's into the growth-rate of crossed and self-fertilized plants.

In the arrangement of the experiment fifteen seeds resulting from each type of fertilization were used; denote these by A-type and B-type seeds. Fifteen pairs of plots, say p_{si} ($s = 1, 2, \ldots 15$, $i = 1, 2$), were chosen and prepared in such a way that the environmental conditions within each pair were as alike as possible. Following the principle of randomization it would then be necessary to determine at random, and for each pair independently, which site should be occupied by A and which by B-type seed.* After the experiment was completed, the grown plants were measured; suppose the character considered (height at given age) had values of a_s and b_s respectively, for the sth pair of plots. Darwin's problem was to determine whether there was any evidence that the type of fertilization affected the vigour of the plant. Statistically, this can be examined by testing the hypothesis, say H_0, that as far as the character measured on the grown plants is concerned, the two samples of seeds have been drawn from identical populations.

* This process of random assignment was not, of course, actually performed by Darwin.

The character in the grown plant depends on (i) the environmental conditions which may be slightly different between plots p_{s1} and p_{s2}, (ii) some quality inherent in the individual seed. We may now imagine in the conceptual field a continued repetition of the experiment, fifteen seeds of each type being randomly selected, fifteen pairs of plots being prepared and a random assignment of the seeds. If then the method of fertilization is unconnected with subsequent growth, a given quality of seed will be as likely to be associated with an A as a B-type seed, and owing to the randomization will be as likely to be associated with the environmental condition of plot p_{s1} as plot p_{s2}. Hence a difference

$$x_s = a_s - b_s$$

of given numerical magnitude will be as likely to be positive as negative. This will be true independently for all fifteen plots.

It follows that the conceptual population of possible experimental results $x_1, x_2, \ldots, x_{15}$ may be divided into an infinite number of subpopulations each defined by a given set of fifteen values of $|x_s|$, and each containing the 2^{15} elements that will be generated by assigning to these numerical values all possible combinations of positive and negative signs. If the hypothesis, H_0, of no differentiation between the A-type and B-type seed populations is true, each of these 2^{15} elements is equally likely to arise.

To construct a test it is now necessary to find a rule, *applicable to every one of these subpopulations*, which will divide the 2^{15} elements into two classes:

(1) a class I containing a proportion P of the elements,

(2) a class II containing a proportion $1-P$ of the elements.

If then we reject the hypothesis, H_0, of no differentiation when the element represented by the fifteen differences $x_1^0, x_2^0, \ldots, x_{15}^0$, actually observed falls into class I, but not otherwise, we may be sure that the risk of rejecting H_0 when it is true is controlled at a value of P: e.g. if $P = \cdot 05$ we should be using what is ordinarily termed a 5 per cent. significance level. The practical question is, of course, how to determine classes I and II. Clearly they should be so determined that if one type of seed in fact produces larger plants than the other, the element represented by the observed differences $x_1^0, \ldots, x_{15}^0$ would be likely to fall into class I, and thus H_0 would, correctly, be rejected. It is seen at once that some consideration of the alternatives to the hypothesis tested is entering into the construction of the test; it has already entered into the design of the experiment since the care taken to make the environmental conditions associated with the pair of plots p_{s1} and p_{s2} similar, was aimed at increasing the chance of detecting a true difference in seed type if one exists.

Fisher's suggestion is to put into class I the $100P$ per cent. of the 2^{15} elements or a number as near that figure as possible for which the fifteen x's have the largest numerical mean value. Thus, for the data of Darwin's experiment, the

values of $x_1^0, x_2^0, \ldots, x_{15}^0$ were: 49, −67, 8, 16, 6, 23, 28, 41, 14, 29, 56, 24, 75, 60, −48, giving a mean of $314/15 = 20{\cdot}9\dot{3}$.

Taking the subpopulation of $2^{15} = 32{,}768$ elements generated by all possible assignments of positive and negative signs to the fifteen values of $|x_s^0|$, Fisher finds that in 1722, or 5·26 per cent., the numerical value of the mean, $\bar{x}$, is greater than the observed value, $20{\cdot}9\dot{3}$. Consequently, the observed result falls just outside the class I associated with a 5 per cent. significance level and we should probably not be prepared to risk rejecting the hypothesis H_0.

The test proposed by Fisher depends upon a particular definition of the class I. It is important to note that this definition is in no sense unique. For example, we could have put into class I the $100P$ per cent. of the 2^{15} elements for which the geometric mean of the fifteen values $(100 + x_s)$ differed most from 100. I do not suggest that this would be a rational classification, but it is worth while reflecting whether, if we choose to use the arithmetic mean as criterion, we are not being influenced, perhaps unconsciously, by

(*a*) the knowledge that if variation is normal, a criterion based on the observed mean difference in samples will be most efficient in detecting a real population difference in seed types;

(*b*) the belief that the characters measured, a_s and b_s, are likely to be approximately normally distributed.

If this is the case, it would seem that the usefulness of the test is in fact dependent on the form of the alternative hypotheses.

Another illustration of the application of this principle of randomization has been recently given by Fisher elsewhere(6). He supposes we have available measures of the stature of a random sample of, say, n Frenchmen and n Englishmen, and wish to test the hypothesis that the mean height of the sampled populations of Frenchmen and Englishmen are identical. Let the observations be written as follows:

$$\begin{cases} \text{Frenchmen } x_1, x_2, \ldots, x_n, \text{ Mean } \bar{x}; \\ \text{Englishmen } y_1, y_2, \ldots, y_n, \text{ Mean } \bar{y}. \end{cases}$$

If the $2n$ observations were written on cards and shuffled without regard to nationality, it would be possible to divide them into a group A and a group B, each containing n cards, in $(2n)!/(n!)^2$ ways. For each way of division we shall have a mean $\bar{a}$ for group A and a mean $\bar{b}$ for group B, giving a difference

$$d = \bar{a} - \bar{b}.$$

Just as in the last example, divide these $(2n)!/(n!)^2$ possible differences into

(1) class I containing the $P\,(2n)!/(n!)^2$ (or a number as near below this as possible) giving largest values of $|\,d\,|$,

(2) class II containing the remaining cases.

Suppose P is chosen to be ·05. Then if the difference $\bar{x} - \bar{y}$ for the observed

French-English subdivision falls into class I, we may reject the hypothesis tested, knowing that the risk we run of rejecting it when it is in fact true is ·05 or less.

This ingenious suggestion of Fisher's leads to the following result: if we adopt the rule wherever a problem of this type arises in our statistical experience, we shall have precise control of the risk of wrong rejection no matter what was the type of variation in the populations sampled.

Of course the procedure needed to determine whether the observed sample falls into class I or class II is very lengthy, unless the samples are very small. I am concerned, however, not with this point, but with the question of whether there is something fundamental about the form of the test suggested, so that it can be used as a standard against which to compare other more expeditious tests, such as Student's. It seems to me that Fisher is overstating the claim of an extremely ingenious device when he writes ((6), p. 59): "Actually, the statistician does not carry out this very simple and very tedious process, but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method." The following example should at any rate help to bring out some points which appear to need careful consideration.

The figures given below represent two samples of seven observations from two populations; they form Experiment I of Table I.

Sample 1. 45, 21, 69, 82, 79, 93, 34. Mean $= \bar{x}_1 = 60{\cdot}43$. Midpoint between extreme values $= m_1 = 57$.

Sample 2. 120, 122, 107, 127, 124, 41, 37. Mean $= \bar{x}_2 = 96{\cdot}86$. Midpoint between extreme values $= m_2 = 82$.

After pooling these fourteen numbers, they can be redivided into two groups A and B, of seven each, in $(14!)/(7!)^2 = 3432$ ways. We may now ask in how many of these ways:

(1) the difference in means of the two groups has an equal or greater negative value than the observed

$$\bar{x}_1 - \bar{x}_2 = 60{\cdot}43 - 96{\cdot}86 = -36{\cdot}43;$$

(2) the difference in midpoints has an equal or greater negative value than the observed

$$m_1 - m_2 = 57 - 82 = -25?$$

After a rather troublesome investigation into the possible arrangements I find the answer to question (1) is 126 out of 3432 or 3·67 per cent., and to question (2) is 45 out of 3432 or 1·31 per cent. It may be said therefore that random assignments of the fourteen numbers into two groups of seven would give (1) as large or a larger numerical value than that observed to the difference in *means* on 7·3 per cent. of occasions, and (2) as large or a larger numerical value to the difference in *midpoints* on 2·6 per cent. of occasions. It follows that in

applying this form of test to the midpoints, we should be more likely to suspect a difference in populations sampled than in applying the test to the means.

Now of course it is quite possible that in individual cases an inferior test may detect a real difference when a better test does not. I give below therefore four further pairs of random samples from the same two populations, as well as the

TABLE I

Experimental Sampling Data

Experiment	I		II		III		IV		V	
Sample	1	2	1	2	1	2	1	2	1	2
	45	120	29	50	14	60	47	60	67	47
	21	122	41	125	70	104	4	90	18	71
	69	107	27	112	32	81	49	84	41	43
	82	127	5	86	79	41	49	100	41	115
	79	124	27	40	87	69	23	93	65	66
	93	41	58	98	25	40	52	32	8	124
	34	37	92	50	2	48	67	98	52	56
Mean	60·43	96·86	39·86	80·14	44·14	63·29	41·57	79·57	41·71	74·57
Midpoint	57·0	82·0	48·5	82·5	44·5	72·0	35·5	66·0	37·5	83·5

TABLE II

Number of pairs of samples, under randomization, having negative values for $\bar{x}_1 - \bar{x}_2$ and $m_1 - m_2$ as great as or greater than the observed pairs

	Mean			Midpoint		
	Greater difference	Equal difference*	Total	Greater difference	Equal difference*	Total
Experiment I	121	5	126	40	5	45
II	56	1	57	44	10	54
III	Over 250	$\geqslant 3$	>253	100	41	141
IV	17	3	20	17	14	31
V	82	2	84	28	25	53

* Including the observed difference itself.

results of applying the two tests. It will be seen that in only one case out of the five does the mean supply stronger evidence of difference than the midpoint. Both these tests are equally valid in the sense that, using either, we can control the error of rejecting the hypothesis that the populations are the same when it is in fact true. In the case taken the population means were at 49·5 and 79·5 respectively and their two standard deviations were the same (= 28·86).

Yet as far as the very limited experimental evidence goes, the midpoint test has been the more effective in detecting the presence of the real difference of 30 units in population means. The reason for this is explained at once when we know that the population distributions were rectangular, e.g.

for population 1 any value of x between 00 and 99 was equally likely to occur; for population 2 any value of x between 30 and 129 was equally likely to occur.*

Since the standard error of the midpoint in samples of n from a rectangular population of standard deviation σ is

$$\sigma_m = \sigma\sqrt{\frac{6}{(n+1)\,(n+2)}},$$

which for $n=7$ is $\cdot 289\sigma$; while for the mean

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}},$$

which for $n=7$ is $\cdot 378\sigma$; we should expect on theoretical grounds that the difference in sample midpoints, rather than in sample means, would be more efficient in detecting real differences in population means. Such a property would certainly appeal to the practical experimenter, were not both tests for other reasons too lengthy to carry out as a common practice.

Now of course in practice it is extremely unlikely that we should deal with variables whose probability distribution is rectangular, but I have introduced these examples because it seems to me to suggest that in problems of this kind it is impossible to make a rational choice between alternative tests unless we introduce some information beyond that contained in the sample data, i.e. some information as to the kind of alternatives with which we are likely to be faced.

If the variation is approximately normal and the standard deviations in the two populations are the same, the advantages of Student's t-test can be expressed in simple terms which appeal to the practical statistician. Its use gives control of the risk of rejecting the hypothesis of "no difference" when it is true, and at the same time makes more probable than does any other test the detection of a real difference in means.† It is certainly possible to claim that these reasons justify its use rather than the relation it bears to the test of Fisher's which I have outlined. It is true that when variation departs from the normal the t-test will not give quite accurate control of the risk of wrong rejection of H_0 (although the error will usually be small), while the test based on randomization will continue to do so. It is in this that the value of the randomization test lies; but as I have pointed out, in so far as this latter test is applied to *means*, it cannot be regarded as unique, and for wide departures from normality it could probably be improved on by use of other central estimates.

* Tippett's Random Sampling Numbers were used; *Tracts for Computers*, No. XV.

† For discussion of this conception see (3) and (4).

3. Randomization applied to the Latin Square

The conceptual model which lies behind the design of the Latin Square experiment leads to the following expression for the yield on the plot in the ith row and jth column receiving the kth treatment:

$$y_{ij(k)} = A + R_i + C_j + T_k + \eta_{ijk}.$$

Here, for a given experiment with an $s \times s$ Latin Square, A, R_i, C_j and T_k ($i = 1, \ldots, s$; $j = 1, \ldots, s$; $k = 1, \ldots, s$) may be regarded as constants and the η's as normally and independently distributed about zero with standard deviation, σ_η. The hypothesis, H_0, which it is generally wished to test is that $T_k = 0$ ($k = 1, \ldots, s$), i.e. that there are no treatment differences.

It has always been recognized, however, that the additive row and column contributions, R_i and C_j, given in this equation cannot provide sufficient elasticity to fit all forms of fertility gradient found in practice. Consequently there is bound to be some correlation among the η's from neighbouring plots, and further the η's may not be normally distributed. In a single experiment it is of course quite impossible to decide whether the s^2 values of η can be reasonably regarded as independent normal deviates. Two lines of procedure seem therefore to have been followed.

In the first place emphasis has been laid on the importance of randomization; in assigning the s treatments to their plots, the particular Latin Square pattern used is chosen at random from the very many possible patterns, say N_s in number. The infinite population of results which can be conceived as obtainable from the experiment, if H_0 is true, may then be divided into an infinite set of subpopulations, each containing a finite number of elements, N_s. Each subpopulation is defined by a set of s^2 yields, y_{ij}, and an element corresponds to a partition of these yields into s treatment groups in accordance with a particular one of the N_s Latin Square patterns. The observed result following from the Latin Square pattern chosen for the experiment represents a single one of these elements.

If now, as far as yield is concerned, the s treatments are identical, it will follow that each of these N_s elements is equally likely to occur owing to the random choice of patterns, even if the η's are not normal or independent. Consequently, as in the previous illustrations, it is only necessary to find a rule, applicable to all sets of s^2 yields, which will enable us to separate from the N_s elements a suitable class I containing a proportion P of them. If this can be done, and the hypothesis of no treatment differences is rejected when the experiment performed gives a result falling into this class, we shall run a risk equal to P of rejecting the hypothesis H_0 when it is true.

Exactly as in the simpler examples, many ways might be found of classifying N_s partitions of the yields, y_{ij}; the choice between them may be influenced by expediency or by the efficiency of the resulting test in detecting the presence of

real treatment differences when they exist. From both points of view it seems reasonable to employ the usual z-criterion, although as soon as we must depart from the original model of the equation above, the fundamental association between sums of squares and normal variation is blurred. Accepting this criterion, class I will consist of the PN_s partitions leading to the largest values of z.

In his paper (7) published on pp. 21–52 above, B. L. Welch has suggested a method of determining approximately the lower limit of z bounding this class and he finds that, if for example $P = \cdot 05$ or $\cdot 01$, this limit does not necessarily correspond exactly to the 5 and 1 per cent. significance levels found from the usual tables of the z probability integral. Where it falls will in fact depend upon the particular set of s^2 yields, y_{ij}. Thus, in one example taken, as few as 2·8 per cent. and in another as few as 2·9 per cent. of the N_s partitions obtained by randomization of yields from a uniformity trial gave values of z above the normal theory 5 per cent. level. This line of approach suggests, therefore, that if we are to obtain a correct probability level for z from the classification of the N_s partitions, it might be necessary to apply a somewhat lengthy procedure to each set of s^2 yields obtained from an experiment.

The second method of attack is one which, while recognizing that the η's may not be exactly independent or normal, asks how far an analysis of uniformity trial data (for which the T_k in the equation are zero) suggests that the distribution of z differs at all seriously from the normal theory form. In this case only a single z is obtained from each experiment, and we are concerned with the distribution of z resulting from experiments which have actually been carried out, rather than that generated hypothetically under randomization when all possible N_s partitions are obtained from the s^2 yields of a single experiment. The investigation carried out by O. Tedin (8) showed that for certain types of Latin Square pattern the distribution of z found in 91 uniformity trials was definitely biased, but for other patterns selected at random this bias was not evident.

It should be noted that even if the assumptions underlying the Latin Square equation were perfectly satisfied, there can be little doubt as a result of B. L. Welch's work that certain sets of plot yields will occur in practice from time to time which, under randomization, will lead to distributions of z differing from normal theory. Some of these distributions, however, would be biased in one way, some in another, so that when they are all combined together the resulting z-distribution should approach that of normal theory. From each randomization set the experimenter is concerned in fact with only one value of z, and this has been selected at random if he has chosen his Latin Square pattern randomly; consequently from the point of view of his long-run experience, the appropriate probability distribution for him to use would appear to be that of normal theory.*

* Possibly we have here another instance of the difference referred to above between regarding a test as giving essentially a rule to be applied and justified by long-run experience, rather than a probability measure associated with an isolated experiment.

On the other hand if the equation fails to represent the situation commonly met with in the field in such a way that there is a general bias in one direction, the resulting under (or over) estimate of significance could be avoided by the lengthy process of referring the observed z in each case to its appropriate randomization distribution.

To throw further light on these points it would certainly seem to be of interest to extend Welch's investigation by applying his results to further uniformity trial data.

The conception of randomization illustrated in the examples given above is both exceedingly suggestive and often practically useful, but perhaps it should be described as a valuable device rather than a fundamental principle. Its adoption, when it can be followed by the calculation necessary to determine what I have described as the class I elements, ensures accuracy in the determination of the probability level of a test criterion, but without the aid of some further principle it cannot help us to decide which of a number of alternative tests to choose. It seems hardly possible to build the methods of statistics into a consistent whole without facing squarely the why of that choice.

REFERENCES

(1) J. Neyman (1934). *J. R. statist. Soc.* xcvii, Appendix, p. 589, and (1937) *Phil. Trans. Roy. Soc.* (at press).
(2) R. A. Fisher (1930). *Proc. Camb. phil. Soc.* xxvi, 528.
(3) J. Neyman and E. S. Pearson (1933). *Philos. Trans.* A, ccxxxi, 289.
(4) —— —— (1936). *Statistical Research Memoirs* (Dept. of Statistics, University College, London), i, 1.
(5) R. A. Fisher (1935). *The Design of Experiments.* Oliver and Boyd.
(6) —— (1936). *J. R. anthrop. Inst.* lxvi, 57.
(7) B. L. Welch (1937). *Biometrika*, xxix, 21.
(8) O. Tedin (1931). *J. agric. Sci.* xxi, 191.

THE DISTRIBUTION OF THE RATIO OF COVARIANCE ESTIMATES IN TWO SAMPLES DRAWN FROM NORMAL BIVARIATE POPULATIONS

By H. O. HIRSCHFELD,

Harper-Adams Agricultural College, Newport, Shropshire

It is well known that the analysis of variance of a single variable necessitates a test of significance, for which Fisher's z-test is the appropriate solution. However, when problems in more than one variable arise, we must consider in addition to the separate variances the question of correlation and covariation. For every kind of analysis the subdivision of the sum of products of the deviations from the respective means into its different components may be performed in exactly the same way as the subdivision of the sum of squares, and what is generally known as an "analysis of variance and covariance" can be worked out easily.

There are three types of problems for which covariance estimates may be used in a test:*

(i) The question whether there is a difference between the regression coefficients of the two normal populations, from each of which we assume one of the samples has been drawn. This test is related to the theory of residual variance in an analysis of variance and covariance.

(ii) The question whether there is a difference between the correlations in the two above populations.

(iii) The question whether there is any difference between corresponding second order parameters.*

In this paper we are mainly interested in question (ii), though the practical example to be considered later is an example for both question (i) and question (ii). We do not deal with question (iii).

A difference between two correlations is most conveniently tested by Fisher's z-transformation of the estimated correlation coefficient. One great advantage of this test is that it is entirely independent of the values of the population variances, i.e. it is valid when nothing whatever is known about the values of the variances.

There are, however, cases in which the estimated variances allow the assumption that corresponding population variances are equal. Such information, however, which may be derived from the variance estimates is purposely ignored

* See, however, the paper by E. S. Pearson and S. S. Wilks(6), where several other problems are dealt with.

when testing correlations estimated from such samples by Fisher's z-transformation.

Assuming, as we may, that corresponding variances of the populations, from which such samples have been drawn, are equal, a difference between their covariances has exactly the same meaning as a difference between their correlations.

This is the type of problem for which a test for a difference between covariances has been developed in this paper with the help of the distribution of the ratio of covariance estimates.*

We shall not enter into a detailed discussion of the appropriateness of this test. It has been said that the distribution of the ratio of covariance estimates is extremely complicated and that it depends on the population parameters. We frankly admit that it essentially depends on the value of the population correlation ρ, and that there is less scope for its common application than for the test of the z-transformation. But it may be regarded as the object of this paper to make a test of this kind at all possible by showing that the distribution of the ratio of covariance estimates may be developed in such a way that its dependence on ρ is of fairly simple character, and by demonstrating how it may be applied to practical examples.

Thus our problem, well defined by the underlying "null-hypothesis", may be stated as follows:

Let x_i, y_i $(i=1, 2, \ldots, n')$ and X_j, Y_j $(j=1, 2, \ldots, N')$ be two samples both drawn from the same normal bivariate population† and let

$$v=(n'-1)^{-1}\sum_{i=1}^{n'}(x_i-\bar{x})(y_i-\bar{y}),$$

$$V=(N'-1)^{-1}\sum_{j=1}^{N'}(X_j-\bar{X})(Y_j-\bar{Y})$$

be the respective estimates of the covariance. We then ask for the chance P (say) of drawing two samples of the above sizes from the population such that the ratio

$$c=v/V$$

is greater (or less) than a certain value c_0 (say). Knowing the value of P for every value of c_0, we then judge the significance of the observed ratio by substituting it for c_0 and comparing the corresponding value of P with the standard levels of ·05 and ·01.

In this paper we shall find the solution of this problem on the assumption that both n' and N', the numbers of items in the samples, are odd numbers. This restriction, unimportant for practical applications, simplifies the mathematical

* In terms of the paper by E. S. Pearson and S. S. Wilks(6) the situation may be characterized by saying that among the set Ω of all pairs of normal populations π_1, π_2 with parameters ξ_{x_1}, ξ_{y_1}, σ_{x_1}, σ_{y_1}, ρ_1 and ξ_{x_2}, ξ_{y_2}, σ_{x_2}, σ_{y_2}, ρ_2 respectively, for which the relations $\sigma_{x_1}=\sigma_{x_2}$, $\sigma_{y_1}=\sigma_{y_2}$ are fulfilled, the hypothesis is tested that in addition $\rho_1=\rho_2$.

† Throughout the paper normal populations are called identical if they have equal variances and correlations.

expression for P considerably, and thus makes it possible to use the latter for a test in practical examples, as will be shown in this paper.

We shall state here the mathematical expression for the chance P, giving the proof in a mathematical appendix. This expression is given in terms of complete and incomplete B-functions, for which tables have been provided by K. Pearson(1). Using his notation we define

$$\mathrm{B}(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}\,dx=(p-1)\,!\,(q-1)\,!/(p+q-1)\,!\,,^*$$

$$\mathrm{B}_X(p,q)=\int_0^X x^{p-1}(1-x)^{q-1}\,dx,$$

$$I_X(p,q)=\mathrm{B}_X(p,q)/\mathrm{B}(p,q).$$

We are now ready to write down:

(*a*) The chance $P^+(\rho)$ that two independent samples, viz.

$$x_i, y_i\,(i=1,2,\ldots,\ n'=2k+3),$$

$$X_j, Y_j\,(j=1,2,\ldots,N'=2K+3),$$

which have been drawn from the same normal population, with correlation coefficient ρ, yield estimates of covariance whose ratio is greater than a certain positive value c_0. We have

$$P^+(\rho)=4^{-k-K-2}k^{-1}K^{-1}\sum_{q=1}^{k+1}\sum_{p=1}^{K+1}I_X(p,q)\frac{2^{p+q}}{\mathrm{B}(k,k+2-q)\,\mathrm{B}(K,K+2-p)}$$
$$\times(1-\rho^2)^{k+K+2}\{(1-\rho)^{-p-q}+(1+\rho)^{-p-q}\},\quad\ldots\ldots(1)$$

where $$X=(K+1)/\{(K+1)+c_0(k+1)\}.$$

(*b*) The chance $P^-(\rho)$ that the above samples yield estimates of covariance whose ratio is smaller than a certain negative value c_0.

$$P^-(\rho)=4^{-k-K-2}k^{-1}K^{-1}\sum_{q=1}^{k+1}\sum_{p=1}^{K+1}2^{p+q}/[\mathrm{B}(k,k+2-q)\times\mathrm{B}(K,K+2-p)]$$
$$\times(1-\rho^2)^{k+K+2}$$
$$\times\{(1-\rho)^{-p}(1+\rho)^{-q}I_{X_1}(p,q)+(1+\rho)^{-p}(1-\rho)^{-q}I_{X_2}(p,q)\},\quad\ldots\ldots(2)$$

where $$X_1=(K+1)(1-\rho)/\{(K+1)(1-\rho)+c_0(k+1)(1+\rho)\},$$
$$X_2=(K+1)(1+\rho)/\{(K+1)(1+\rho)+c_0(k+1)(1-\rho)\}.$$

The expressions for $P(\rho)$ are thus *finite* weighted sums of incomplete B-functions, the weights being complete B-functions and simple polynomials in ρ^2. For $\rho=0$ the chances (1) and (2) are identical, i.e. the distribution is symmetrical in c_0. As ρ^2 tends to 1 (i.e. as ρ tends to $+1$ or to -1) equation (1) approaches the simplified form

$$P^+(\pm 1)=I_X(k+1,K+1),\quad\ldots\ldots(1')$$

* For integral values of p and q.

which is the representation of Fisher's z-test in terms of Pearson's incomplete B-functions. Thus, whenever it is known that there is a very high correlation between the variates x and y, covariance estimates might approximately be tested like mean squares by Fisher's z-test.

On the other hand, since expression (2) approaches 0 as $\rho^2 \to 1$, negative values of c_0 are less and less likely to occur as ρ^2 increases and there will be but few problems with reasonably correlated variates, in which (2) has to be used for a test.

From formulae (1) and (2) it is clear that the chance P essentially depends on the correlation coefficient ρ of the population, which is unknown. This property, disadvantageous for a test of significance, is obviously characteristic for the nature of our problem. To demonstrate this dependence eleven different ρ-values covering the interval $-1 \leqq \rho \leqq +1$ were chosen, viz. the values $\rho = 0, \pm \cdot 2, \pm \cdot 4, \pm \cdot 6, \pm \cdot 8, \pm 1$, and for these values were calculated the chances $P(\rho)$ (given by equation (1)) of obtaining two samples of 15 ($k = K = 6$) whose ratio of covariance estimates is greater than 74/26.*

The result is given below in Table I.

TABLE I

Giving the chance $P(\rho)$ of drawing two random samples of 15 from a normal population with correlation coefficient ρ such that the ratio of their covariance estimates is greater than 74/26

$\rho =$	0	$\pm \cdot 2$	$\pm \cdot 4$	$\pm \cdot 6$	$\pm \cdot 8$	± 1
$P(\rho) =$	·110	·128	·135	·097	·054	·030

From Table I it is obvious that there is no hope of approximating to our distribution (or to a transformation of it) by a normal curve (or any suitable distribution function) *independent of* ρ. Therefore, in testing significance we must admit all possible values of ρ, as we have started to do in Table I. Thus three different types of results may occur:

(*a*) For all values of ρ the P-values are smaller than ·05 (significant at 5 per cent.).

(*b*) For all values of ρ the P-values are greater than ·05 (insignificant at 5 per cent.).

(*c*) For some values of ρ the P-values are smaller than ·05, for other ρ-values the P-values are greater than ·05. The former ρ-values will cover an interval (or a set of intervals) on $-1 \leqq \rho \leqq +1$, which may be called I_1, the latter ρ-values will cover the remaining part of the range $-1 \leqq \rho \leqq +1$, which may be called I_2.

Table I shows an example of type (*c*), I_1 being (approximately) the intervals $\cdot 8 \leqq |\rho| \leqq 1$, I_2 being the interval $0 \leqq |\rho| \leqq \cdot 8$. To complete the test in cases like

* The value 74/26 has been chosen in connexion with a practical example to be discussed later.

this, the best estimate r of the correlation coefficient ρ has to be calculated from the samples, and the deviation of this observed value r from the nearest value in I_2, ρ' (say), taken as population value, has to be examined. In case this deviation is insignificant (at 5 per cent.), the whole test returns an insignificant result. For it has failed to disprove the hypothesis that both samples have been drawn from the same population with correlation coefficient ρ'. If, on the other hand, the deviation is significant, so is our original test. For, whatever the value of ρ may be, the original "null-hypothesis" has been disproved.

This latter part of the test is most easily worked out by Fisher's approximate method ((2), § 35) or by entering the exact distribution of r with ρ' as population value,* e.g. (1), p. xxxviii.

The first and main part of our test, however, consists in calculating the above chance $P(\rho)$ (see equations (1) and (2)), or rather in finding among all ρ-values, for which $P(\rho) \geqq \cdot 05$ (i.e. among all ρ-values in I_2, if any) that value ρ', which is nearest to our observed correlation coefficient.

To facilitate this, trivariate tables for the 5 per cent. points of the distribution of c (or a suitable transformation of it) would have to be worked out. The best arrangement of these would be in such a way, that two-way tables with the number of items in the larger sample as row-headings and (say) twenty different, positive ρ-values as column headings should proceed in pages with the number of items in the smaller sample.†

But even without the aid of such tables there is a method of working out the test for practical examples with the help of Pearson's tables of the incomplete B-function, and the calculations for such an example are shown below. Since Pearson's tables have been prepared so as to answer various other purposes, much calculation work is still left to be done, when applying them to our test. Though a further table (Table II) has been prepared, which facilitates the work considerably, the following method, unless both samples are very small, is still too laborious to become a common statistical practice.

PRACTICAL EXAMPLE

In a Cambridge nutrition experiment on pigs(3), among many other post-slaughter results the "mean back fat" (x) and the "percentage fat from back to belly" (y) were measured for 15 hogs and 15 gilts. One problem was to see how far mean back fat, which measurement is taken without great difficulty, provides a fair estimate of the percentage fat from back to belly and thus may be used for grading purposes.

We do not reproduce the $30+30$ measurements here, but on examination it would be noted that the relationship was fairly marked for hogs but not so for

* In this case it might become necessary to perform the test for two ρ' values, viz. the nearest ρ-value in I_2 with $\rho < r$ (if any) and the nearest ρ-value in I_2 with $\rho > r$ (if any).

† The author regrets that, at the moment, he is unable to undertake this work, since he has only an adding machine at his disposal.

gilts. This sex difference is confirmed by the values of the respective sums of squares and products given below:

	(x^2)	(xy)	(y^2)	
Hogs	1·0975	10·488	236·06	.....(3).
Gilts	0·6544	3·527	302·58	

Testing the significance of a correlation between y and x (or of a regression of y on x) for hogs and gilts separately a highly significant result is obtained for hogs whilst the gilts regression is quite insignificant. Nevertheless, we shall see that the difference between these relationships cannot be regarded as being significant. We shall test both the difference between correlation estimates (with the help of Fisher's z-transformation) and the difference between regression estimates (by the t-test), for in this example both questions are of interest. Finally, since variance estimates allow the assumption of equal population variances for hogs and gilts, we shall compare these tests with the test for the ratio of covariance estimates.

Let us start with the t-test. Doing this we obtain for b_H and b_G the respective values

$$b_H = 9{\cdot}56, \quad b_G = 5{\cdot}39.$$

Furthermore, the estimated S.D. of the difference $b_H - b_G$ has to be calculated in the usual way, the work being shown below:

Residual sum of squares (hogs) =	135·83
„ „ (gilts) =	283·57
$26 \times s^2 =$	419·40
$s^2 =$	16·131
$s^2/1{\cdot}0975 =$	14·70
$s^2/\ {\cdot}6544 =$	24·65
	39·35
S.D. of $(b_H - b_G) = \sqrt{39{\cdot}35}$	
$t \sim$	·665

The t-test returns an altogether insignificant difference; the chance of obtaining a difference equal to or larger than that observed being greater than 0·5.

Next we consider Fisher's test for the difference $z_1 - z_2$, i.e. the difference between the z-transformations of the estimated correlations r_1 and r_2. We obtain (approx.):

$$\left.\begin{array}{ll} r_1 = {\cdot}65, & z_1 = {\cdot}78 \\ r_2 = {\cdot}25, & z_2 = {\cdot}25 \end{array}\right\} \text{ diff.} = {\cdot}53;$$

S.D. of diff. $= {\cdot}408;$

normal deviate $= 1{\cdot}3.$

Again we obtain an insignificant result. The chance of obtaining a difference between correlation estimates larger than that observed is about 0·2.

The ratio of the covariance estimates, however, viz. that of the hogs divided by that of the gilts, is nearly 3, and since we know that our test approaches the z-test as ρ^2 approaches 1, this ratio will be significant for large ρ^2. The actual result of our test is summarized in Table I showing significance at 5 per cent. for $\cdot 8 \leqq |\rho| \leqq 1$ (approx.) and insignificance for $0 \leqq |\rho| \leqq \cdot 8$.*

Calculating now from equation (3) the best estimate of our correlation coefficient common to hogs and gilts, we obtain

$$r = \cdot 46.$$

This value lies right inside our I_2 interval and thus, having no deviation from the "nearest" ρ-value in I_2, returns our ratio c_0 as insignificant.

Though we obtain the same result as with the usual tests, it is obvious that in this example our test is more sensitive to the difference between the above covariance estimates. For our largest P-value is about 0·14 whilst the regression test yielded a P greater than 0·5 and Fisher's test of the z-transformation a P-value of about 0·2.

An explanation of the calculations, on which Table I is based, follows: We have to compute the value of $P^+(\rho)$ (given by equation (1)) for $k = K = 6$, $\rho = 0$, $\pm \cdot 2$, $\pm \cdot 4$, $\pm \cdot 6$, $\pm \cdot 8$, ± 1 and $c_0 = 74/26$.†

We first transform formula (1) into a form which is more suitable for its computation, whenever $k = K$.

Introducing the abbreviation

$$C(p,q;\rho) = 4^{-k-K-2}(kK)^{-1} \times 2^{p+q}[\mathrm{B}(k, k+2-q)\,\mathrm{B}(K, K+2-p)]^{-1} \times (1-\rho^2)^{k+K+2}\{(1-\rho)^{-p-q} + (1+\rho)^{-p-q}\},$$

we have

$$P^+(\rho) = \sum_{q=1}^{k+1}\sum_{p=1}^{K+1} I_X(p,q)\,C(p,q;\rho).$$

Now, since the I-function is only tabulated for $p \geqq q$, we write

$$P^+(\rho) = \sum_{q=1}^{k+1}\sum_{p=q}^{K+1} I_X(p,q)\,C(p,q;\rho) + \sum_{q=2}^{k+1}\sum_{p=1}^{q-1} I_X(p,q)\,C(p,q;\rho) \quad \ldots\ldots(4).$$

But since

$$I_X(p,q) = 1 - I_{1-X}(q,p),\ k = K \text{ and thus } C(p,q;\rho) = C(q,p;\rho),$$

we may write instead of equation (4)

$$P^+(\rho) = \sum_{q=1}^{k+1}\sum_{p=q}^{k+1} I_X(p,q)\,C(p,q;\rho) + \sum_{q=1}^{k}\sum_{p=q+1}^{k+1}\{1 - I_{1-X}(p,q)\}\,C(p,q;\rho),$$

* The value 74/26 shown in Table I is slightly smaller than the actual value of c_0 observed, viz. 2·974. This has been done to simplify the work and will be explained later.

† See the footnote to p. 68.

and finally we arrive at

$$P^{+}(\rho)=k^{-2}4^{-2k-2}(1-\rho^2)^{2k+2}\left[\sum_{q=1}^{k+1}\frac{1}{\mathrm{B}(k,k+2-q)}\sum_{p=q}^{k+1}I_X(p,q)\frac{2^{p+q}}{\mathrm{B}(k,k+2-p)}\right.$$

$$\left.\times Q(\rho^2,p+q)+\sum_{q=1}^{k}\frac{1}{\mathrm{B}(k,k+2-q)}\sum_{p=q+1}^{k+1}[1-I_{1-X}(p,q)]\frac{2^{p+q}}{\mathrm{B}(k,k+2-p)}Q(\rho^2,p+q)\right], \quad \ldots\ldots(5)$$

where $$Q(\rho^2,p+q)=\{(1-\rho)^{-p-q}+(1+\rho)^{-p-q}\}.$$

These two sums are now most conveniently worked out together. The first step consists in preparing a table of the values

$$(1-\rho)^{-p-q}+(1+\rho)^{-p-q}.$$

This table, because independent of c_0, may be used by the reader for similar tests, provided the number of observations, when added for the two samples, is not greater than 30. The accuracy which is required for the entries of this table depends on the size of the sample, and (especially for high values of ρ) increases considerably as $p+q$ increases. Therefore a 5-figure table has been prepared rather than giving entries accurate up to a certain decimal.

TABLE II

Values of $(1-\rho)^{-p-q}+(1+\rho)^{-p-q}$

$(p+q)$	(*b*) $\rho=\cdot2$	(*c*) $\rho=\cdot4$	(*d*) $\rho=\cdot6$	(*e*) $\rho=\cdot8$
2	2·2569	3·2880	6·6406	$2\cdot5309\times10$
3	2·5318	4·9941	$1\cdot5869\times10$	$1\cdot2517\times10^{2}$
4	2·9237	7·9764	$3\cdot9215^{+}\times10$	$6\cdot2510\times10^{2}$
5	3·4536	$1\cdot3046\times10$	$9\cdot7752\times10$	$3\cdot1251\times10^{3}$
6	4·1496	$2\cdot1566\times10$	$2\cdot4420\times10^{2}$	$1\cdot5625\times10^{4}$
7	$5\cdot0475^{-}$	$3\cdot5817\times10$	$6\cdot1039\times10^{2}$	5^{p+q}
8	6·1930	$5\cdot9605^{+}\times10$	$1\cdot5259\times10^{3}$	
9	7·6444	$9\cdot9277\times10$	$3\cdot8147\times10^{3}$	
10	9·4747	$1\cdot6542\times10^{2}$	$9\cdot5368\times10^{3}$	
11	$1\cdot1776\times10$	$2\cdot7566\times10^{2}$	$2\cdot3842\times10^{4}$	
12	$1\cdot4664\times10$	$4\cdot5941\times10^{2}$	$5\cdot9605^{-}\times10^{4}$	
13	$1\cdot8283\times10$	$7\cdot6567\times10^{2}$	$1\cdot4901\times10^{5}$	
14	$2\cdot2815^{+}\times10$	$1\cdot2761\times10^{3}$	$3\cdot7253\times10^{5}$	

Next we write down the values of $[\mathrm{B}(6,8-q)]^{-1}$ for $q=1,\ \ldots,\ 7$, which are easily obtained from their definition, viz.

$$[\mathrm{B}(6,8-q)]^{-1}=(13-q)!/(7-q)!\,5!$$

Finally, we prepare with the help of Pearson's tables of the incomplete B-function a table of the values

$$2^{p+q}[\mathrm{B}(6,8-p)]^{-1}\{I_{\cdot26}(p,q)+1-I_{\cdot74}(p,q)\}\text{ for }p>q.$$

TABLE III

Values of $1/B(6, 8-q)$

q	1	2	3	4	5	6	7
$1/B(6, 8-q)$	5544	2772	1260	504	168	42	6

In choosing ·26 for X we have substituted for c_0 a value which is slightly smaller than the ratio actually observed, in order to coincide with an entry in Pearson's table. This we have done to save the reader the work of about 56 interpolations. The diagonal entries of the table are the values $2^{2p}[B(6, 8-p)]^{-1} \times I_{\cdot 26}(p, p)$, since these are required for the further calculations.

TABLE IV

Values of $\frac{2^{p+q}}{B(6, 8-p)}\{I_{\cdot 26}(p, q) + 1 - I_{\cdot 74}(p, q)\}$ *for* $p > q$ *and* $\frac{2^{p+q}}{B(6, 8-p)} I_{\cdot 26}(p, p)$ for $p = q$

p \ q	1	2	3	4	5	6	7	(*)
1	5,765·8							
2	11,531·5	7,435·5						
3	12,345·0	13,519·1	9,220·6					
4	11,365·5	13,107·9	14,752·9	10,342·0				
5	8,378·9	10,649·1	12,040·7	13,789·4	9,829·0			
6	4,494·9	6,250·3	7,564·2	8,478·3	9,829·0	7,085·8		
7	1,349·5	2,020·9	2,636·2	3,101·5	3,461·4	4,049·0	2,943·2	
(*a*) $\rho = 0$	$11{,}046 \times 10$	$10{,}597 \times 10$	$9{,}243 \times 10$	$7{,}142 \times 10$	$4{,}624 \times 10$	$2{,}227 \times 10$	5,886	$1{,}067 \times 10^6$
(*b*) $\rho = \cdot 2$	$1{,}834 \times 10^2$	$2{,}307 \times 10^2$	$2{,}701 \times 10^2$	$2{,}863 \times 10^2$	260×10^3	178×10^3	67×10^3	219×10^7
(*c*) $\rho = \cdot 4$	75×10^4	$1{,}473 \times 10^3$	$2{,}632 \times 10^3$	$4{,}243 \times 10^3$	593×10^4	636×10^4	38×10^5	150×10^8
(*d*) $\rho = \cdot 6$	87×10^5	$2{,}856 \times 10^4$	836×10^5	$2{,}232 \times 10^5$	534×10^6	$1{,}026 \times 10^6$	110×10^7	48×10^{10}
(*e*) $\rho = \cdot 8$	105×10^7	75×10^8	465×10^8	265×10^9	$1{,}421 \times 10^9$	$6{,}673 \times 10^9$	$1{,}796 \times 10^{10}$	85×10^{13}

The rest of the work is obvious from formula (5) and is shown in the above table.†

In rows (*a*), (*b*), (*c*), (*d*) and (*e*) are given weighted totals of the respective columns in the upper part of Table IV. To calculate the weighted column totals which are shown in row (*b*) (say) we multiply the entry in the pth row and in the qth column of Table IV by that entry of Table II which is shown in the $(p+q)$th row and in the column (*b*). Then these products are summed to yield the weighted column totals in row (*b*) of Table IV. Similarly, for the rows (*c*), (*d*) and (*e*). The entries of row (*a*) are simply obtained by doubling the column totals of Table IV.

Finally the qth entry of each of the rows (*a*), (*b*), (*c*), (*d*) and (*e*) in Table IV is multiplied by the qth entry of Table III and these products summed for each row,

† The great loss of accuracy is due to the small scope of an ordinary adding machine, with which the work had to be done.

the respective sums of products being shown in column (*). If these sums of products are divided by the respective values of $36 \times 4^{14} \times (1-\rho^2)^{-14}$, the chances $P^+(\rho)$ of Table I are obtained.

My heartiest thanks are due to Dr J. Wishart for his help and advice throughout this work and to Mr F. J. Dudley for his help in preparing Table II.

APPENDIX

Derivation of the distribution function for the ratio of covariance estimates

(i) In this part we shall give a mathematical proof of the formulae (1) and (2), which have been used for a practical test in the preceding part of the paper. Incidentally we shall derive the theory for a more general problem and point out further properties of our distribution, which are analogous to the well-known behaviour of the z-distribution.

The problem may be stated in its generalized form straightaway:

Let $x_i, y_i \quad (i = 1, 2, \ldots, n' = 2k+3)$

and $X_j, Y_j \quad (j = 1, 2, \ldots, N' = 2K+3)$

be two (independent) random samples drawn from the populations

$$f(x,y) = [(\alpha\beta - \nu^2)^{\frac{1}{2}}/\pi] \times \exp\{-(\alpha x^2 + 2\nu xy + \beta y^2)\} \quad \ldots\ldots(6)$$

and $$F(X,Y) = [(\mathrm{AB} - \mathrm{N}^2)^{\frac{1}{2}}/\pi] \times \exp\{-(\mathrm{A}X^2 + 2\mathrm{N}XY + \mathrm{B}Y^2)\} \quad \ldots\ldots(7)$$

respectively.†

It is then required to obtain the distribution function for the ratio of covariance estimates

$$c = [(N'-1)/(n'-1)] \times \sum_{i=1}^{n'} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{j=1}^{N'} (X_j - \overline{X})(Y_j - \overline{Y}),$$

where by $\bar{x}$, $\bar{y}$, $\overline{X}$, $\overline{Y}$ we denote the respective arithmetic means of the samples, viz.

$$\frac{1}{n'}\sum_{i=1}^{n'} x_i, \quad \frac{1}{n'}\sum_{i=1}^{n'} y_i, \quad \frac{1}{N'}\sum_{j=1}^{N'} X_j, \quad \frac{1}{N'}\sum_{j=1}^{N'} Y_j.$$

In the course of the proof, the ratio of the "sums of products",

$$w = \sum_{i=1}^{n'} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{j=1}^{N'} (X_j - \overline{X})(Y_j - \overline{Y}), \quad \ldots\ldots(8)$$

will turn out to be a more convenient statistic than c itself.

† Instead of the parameters α, β, ν the quantities σ_1, σ_2, ρ (i.e. the standard deviations of x and of y and the correlation between x and y) are more commonly used to represent a normal population. In terms of σ_1, σ_2, ρ the parameters α, β, ν are defined by

$$\alpha^{-1} = 2\sigma_1^2(1-\rho^2), \quad \beta^{-1} = 2\sigma_2^2(1-\rho^2), \quad \nu = -\rho/\{2\sigma_1\sigma_2(1-\rho^2)\}.$$

(ii) Our first step is to show that the distribution of the sum of products

$$u=\sum_{i=1}^{n'}(x_i-\bar{x})(y_i-\bar{y}) \qquad \text{.....(9)}$$

may be expressed as a finite sum of elementary functions, provided n', the number of items in the sample, is odd. If we introduce new variates ξ, η by the linear transformation

$$\left.\begin{aligned} x&=\sqrt{\beta/\alpha}\,\xi+\eta \\ y&=\xi-\sqrt{\alpha/\beta}\,\eta \end{aligned}\right\} \qquad \text{.....(10)}$$

then by substituting (10) in (6) we obtain for the distribution of ξ, η

$$[2(\alpha\beta-\nu^2)^{\frac{1}{2}}/\pi]\times\exp\{-(2\beta+2\nu\sqrt{\beta/\alpha})\,\xi^2-(2\alpha-2\nu\sqrt{\alpha/\beta})\,\eta^2\}. \qquad \text{.....(11)}$$

Now to any sample of n' pairs x_i, y_i there will correspond a sample of n pairs ξ_i, η_i obtained by the converse of (10). Furthermore, obviously

$$u=\sqrt{\beta/\alpha}\,\gamma_{11}-\sqrt{\alpha/\beta}\,\gamma_{22},$$

$$\text{if} \qquad \gamma_{11}=\sum_{i=1}^{n'}(\xi_i-\bar{\xi})^2, \quad \gamma_{22}=\sum_{i=1}^{n'}(\eta_i-\bar{\eta})^2, \quad \bar{\xi}=\left(\frac{1}{n'}\right)\sum_{i=1}^{n'}\xi_i, \quad \bar{\eta}=\left(\frac{1}{n'}\right)\sum_{i=1}^{n'}\eta_i. \qquad \text{.....(12)}$$

From equation (11) it is obvious that the variates ξ, η are independently distributed. Hence the joint distribution of γ_{11}, γ_{22} (see e.g. (4)) is given by

$$\psi(\gamma_{11},\gamma_{22})=4^{k+1}(\alpha\beta-\nu^2)^{k+1}(k\,!)^{-2}\gamma_{11}^k\gamma_{22}^k \times\exp\{-(2\beta+2\nu\sqrt{\beta/\alpha})\,\gamma_{11}-(2\alpha-2\nu\sqrt{\alpha/\beta})\,\gamma_{22}\},$$

where $k=(n'-3)/2$ is an integer $\geqslant 1$. Thus by equation (12) we arrive at the joint distribution of u and γ_{22}, viz.

$$\chi(u,\gamma_{22})=4^{k+1}(\alpha\beta-\nu^2)^{k+1}(k\,!)^{-2}(\sqrt{\alpha/\beta})^{k+1}\exp\{-(2\sqrt{\alpha\beta}+2\nu)\,u\} \times(u+\sqrt{\alpha/\beta}\,\gamma_{22})^k\gamma_{22}^k\exp\{-4\alpha\gamma_{22}\}. \qquad \text{.....(13)}$$

To obtain the distribution function $g(u)$ (say) of the sum of products we have to integrate $\chi(u,\gamma_{22})$ (see equation (13)) over the range of γ_{22}. This range, however, depends on u. For since the range of the variates γ_{11}, γ_{22} is given by

$$0\leqq\gamma_{11}<\infty, \quad 0\leqq\gamma_{22}<\infty,$$

the range for u, γ_{22} must be

$$0\leqq\gamma_{22}<\infty, \quad -u\sqrt{\beta/\alpha}\leqq\gamma_{22}, \quad -\infty<u<+\infty.$$

We first consider the case $u\geqq 0$.

In this case we have to integrate $\chi(u,\gamma_{22})$ for γ_{22} ranging from 0 to $+\infty$. To do this

we apply partial integration $(2k+1)$ times, integrating $\exp\{-4\alpha\gamma_{22}\}$, differentiating the powers of γ_{22} and collecting the terms at $\gamma_{22}=0$. We thus obtain

$$g^+(u)=4^{k+1}(\alpha\beta-\nu^2)^{k+1}(k\,!)^{-2}(\sqrt{\alpha/\beta})^{k+1}\exp\{-(2\sqrt{\alpha\beta}+2\nu)u\}$$
$$\times\sum_{\tau=0}^{k}\frac{(k+\tau)!\,k!}{(k-\tau)!\,\tau!}(\sqrt{\alpha/\beta})^{\tau}u^{k-\tau}(4\alpha)^{-k-\tau-1},$$

which may be written as

$$g^+(u)=(\alpha\beta-\nu^2)^{k+1}\exp\{-(2\sqrt{\alpha\beta}+2\nu)u\}(\sqrt{\alpha\beta})^{-k-1}$$
$$\times\sum_{\tau=0}^{k}\frac{(k+\tau)!\,(4\sqrt{\alpha\beta})^{-\tau}}{k!\,\tau!\,(k-\tau)!}u^{k-\tau}. \quad \ldots\ldots(14)$$

We next consider the case $u\leqq 0$.

Now we have to integrate for γ_{22} ranging from $-u\sqrt{\beta/\alpha}$ to $+\infty$. Integrating by parts as above and collecting the terms at $\gamma_{22}=-u\sqrt{\beta/\alpha}$ we have

$$g^-(u)=(\alpha\beta-\nu^2)^{k+1}\exp\{(2\sqrt{\alpha\beta}-2\nu)u\}(\sqrt{\alpha\beta})^{-k-1}$$
$$\times\sum_{\tau=0}^{k}\frac{(k+\tau)!\,(4\sqrt{\alpha\beta})^{-\tau}}{k!\,\tau!\,(k-\tau)!}(-u)^{k-\tau}. \quad \ldots\ldots(15)$$

The required distribution function for u is thus given by

$$g(u)=\begin{cases}g^+(u) & \text{for} \quad u\geqq 0 \quad \text{(see equation (14))}\\ g^-(u) & \text{for} \quad u\leqq 0 \quad \text{(see equation (15))}\end{cases}. \quad \ldots\ldots(16)$$

Comparing these with the "Bessel-function distribution" of $2\sqrt{\alpha\beta}\,u$ [5], [4], we incidentally have proved the well-known fact that the Bessel-function $K_\nu(x)$, for fractional ν, can be expressed by elementary functions in a finite form.

(iii) It is now easy to see how the distribution of the ratio of the two sums of products

$$w=\frac{u}{U}=c\,[(k+1)/(K+1)]$$

can be derived from equations (14) and (15) by elementary integrations. For if we consider the sum of products

$$U=\sum_{j=1}^{2K+3}(X_j-\overline{X})(Y_j-\overline{Y}),$$

the sample of which has been drawn from the population (7), then the distribution of U (say $G(U)$) is obtained by replacing in equations (14), (15) and (16) the letters u, α, β, ν, g, $k=(n'-3)/2$ by U, A, B, N, G, $K=(N'-3)/2$, respectively. Hence, independence of the samples (x_i, y_i) and (X_j, Y_j) being assumed, the joint distribution of u and U is equal to $g(u)\times G(U)$ and thus the distribution function of w (say $\Phi(w)$) is given by

$$\Phi(w)=\int_{-\infty}^{+\infty}g(wU)\,G(U)\,|\,U\,|\,dU, \quad \ldots\ldots(17)$$

since the modulus of the Jacobian of the transformation

$$u = wU, \quad U = U$$

is simply equal to $|U|$. To work out equation (17) we first consider the case

$$w \geqq 0.$$

We now may write

$$\Phi^+(w) = -\int_{-\infty}^{0} g^-(wU)\, G^-(U)\, U\, dU + \int_0^{\infty} g^+(wU)\, G^+(U)\, U\, dU$$
$$= T_1(w) \qquad\qquad + T_2(w) \text{ (say)},$$

and start to consider $T_2(w)$. Now by equation (14) plainly

$$T_2(w) = (\alpha\beta - \nu^2)^{k+1} (\sqrt{\alpha\beta})^{-k-1} (AB - N^2)^{k+1} (\sqrt{AB})^{-k-1} (k!\,K!)^{-1}$$
$$\times \sum_{\tau=0}^{k} \frac{(k+\tau)!}{(k-\tau)!\,\tau!} \sum_{\mu=0}^{K} \frac{(K+\mu)!}{(K-\mu)!\,\mu!} (4\sqrt{\alpha\beta})^{-\tau} (4\sqrt{AB})^{-\mu}$$
$$\times w^{k-\tau} \int_0^{\infty} U^{k+K-\tau-\mu+1} \exp\{-2[(\sqrt{\alpha\beta}+\nu)w + (\sqrt{AB}+N)]U\}\, dU,$$

where the last integral, by partial integration, is seen to be equal to

$$(k+K-\tau-\mu+1)!\,[2\{(\sqrt{\alpha\beta}+\nu)w + (\sqrt{AB}+N)\}]^{-k-K+\tau+\mu-2}.$$

If then $T_1(w)$ is worked out in the same way with the help of equation (15) we finally have

$$\Phi^+(w) = [(\alpha\beta - \nu^2)/\sqrt{\alpha\beta}]^{k+1}\, [(AB - N^2)/\sqrt{AB}]^{K+1}\, (k!\,K!)^{-1}$$
$$\times \sum_{\tau=0}^{k} \frac{(k+\tau)!}{(k-\tau)!\,\tau!} \sum_{\mu=0}^{K} \frac{(K+\mu)!}{(K-\mu)!\,\mu!} (2\sqrt{\alpha\beta})^{-\tau} (2\sqrt{AB})^{-\mu}\, 2^{-k-K-2}$$
$$\times (k+K-\tau-\mu+1)!\, w^{k-\tau} \{[(\sqrt{\alpha\beta}+\nu)w + (\sqrt{AB}+N)]^{-k-K+\tau+\mu-2}$$
$$+ [(\sqrt{\alpha\beta}-\nu)w + (\sqrt{AB}-N)]^{-k-K+\tau+\mu-2}\}, \quad \ldots\ldots(18)$$

or, introducing $i = k - \tau$ and $j = K - \mu$ as summation indices,

$$\Phi^+(w) = [(\alpha\beta - \nu^2)/\sqrt{\alpha\beta}]^{k+1}\, [(AB - N^2)/\sqrt{AB}]^{K+1}\, (k!\,K!)^{-1}$$
$$\times \sum_{i=0}^{k} \frac{(2k-i)!}{i!\,(k-i)!} \sum_{j=0}^{K} \frac{(2K-j)!}{j!\,(K-j)!} \frac{(2\sqrt{\alpha\beta})^{-k+i}\, (2\sqrt{AB})^{-K+j}}{2^{k+K+2}} (i+j+1)!$$
$$\times w^i \{[(\sqrt{\alpha\beta}+\nu)w + (\sqrt{AB}+N)]^{-i-j-2} + [(\sqrt{\alpha\beta}-\nu)w + (\sqrt{AB}-N)]^{-i-j-2}\}.$$
$$\ldots\ldots(19)$$

Turning now to the case $w \leqq 0$

we obtain by the same argument as above for

$$\Phi^-(w) = \int_{-\infty}^{0} g^+(wU)\, G^-(U)\, |U|\, dU + \int_0^{\infty} g^-(wU)\, G^+(U)\, U\, dU$$

the sum

$$\Phi^-(w) = [(\alpha\beta - \nu^2)/\sqrt{\alpha\beta}]^{k+1}\,[(AB - N^2)/\sqrt{AB}]^{K+1}\,(k!\,K!)^{-1}$$
$$\times \sum_{i=0}^{k} \frac{(2k-i)!}{i!\,(k-i)!} \sum_{j=0}^{K} \frac{(2K-j)!}{j!\,(K-j)!} \frac{(2\sqrt{\alpha\beta})^{-k+i}\,(2\sqrt{AB})^{-K+j}}{2^{k+K+2}}\,(i+j+1)!$$
$$\times (-w)^i\,\{[(\sqrt{\alpha\beta}+\nu)(-w)+(\sqrt{AB}-N)]^{-i-j-2}$$
$$+[(\sqrt{\alpha\beta}-\nu)(-w)+(\sqrt{AB}+N)]^{-i-j-2}\}. \quad \ldots\ldots(20)$$

(iv) To use equations (19).and (20) for a test of significance we confine ourselves to the most important special case, viz. we assume that the populations (6) and (7) are identical, whence

$$\alpha = A, \quad \beta = B, \quad \nu = N.^* \qquad \ldots\ldots(21)$$

We first consider the distribution $\Phi(w)$ for

$$w \geqq 0.$$

Using equation (21) and introducing the correlation coefficient

$$\rho = -\nu/\sqrt{\alpha\beta},$$

equation (19) may be written as

$$\Phi^+(w) = \sum_{i=0}^{k} \sum_{j=0}^{K} \frac{(2k-i)!}{(k-i)!\,k!} \frac{(2K-j)!}{(K-j)!\,K!} \frac{(i+j+1)!}{i!\,j!} \frac{w^i}{(1+w)^{i+j+2}}$$
$$\times 2^{-2k-2K+i+j-2}\,(1-\rho^2)^{k+K-i-j}\,\{(1-\rho)^{i+j+2}+(1+\rho)^{i+j+2}\}. \quad \ldots\ldots(22)$$

Thus $\Phi^+(w)$ is seen to depend on $k = (n'-3)/2$, $K = (N'-3)/2$ and ρ^2. Furthermore, as $\rho^2 \to 1$,

$$\Phi^+(w) \to \frac{(k+K+1)!}{k!\,K!} \frac{w^k}{(1+w)^{k+K+2}},$$

and this is the distribution function for $w = [(k+1)/(K+1)]\,e^{2z}$.

To prove formula (1) we have to determine the probability integral

$$P^+ = \int_W^\infty \Phi^+(w)\,dw \quad \text{for any } W \geqq 0.$$

But since
$$\int_W^\infty \frac{w^i}{(1+w)^{i+j+2}}\,dw = -\int_{1/(1+W)}^{0} \omega^{i+j+2}\,(1-\omega)^i\,\omega^{-i-2}\,d\omega$$
$$= \int_0^{1/(1+W)} \omega^j\,(1-\omega)^i\,d\omega,$$

we see from equation (22) that we can express our probability integral with the help of the incomplete B-functions. Using the notation of p. 67 we obtain

$$P^+ = \int_W^\infty \Phi^+(w)\,dw = 4^{-k-K-2}\,k^{-1}\,K^{-1}$$
$$\times \sum_{q=1}^{k+1} \sum_{p=1}^{K+1} I_{[1/1+W]}(p,q)\,\frac{2^{q+p}}{B(k,k+2-q)\,B(K,K+2-p)}$$
$$\times (1-\rho^2)^{k+K+2-p-q}\,\{(1-\rho)^{p+q}+(1+\rho)^{p+q}\}, \qquad \ldots\ldots(23)$$

which is equivalent to formula (1).

* In order to derive formulae (1) and (2) it would be sufficient to assume that for the populations (6) and (7) the equations $\nu = N$ and $\alpha\beta = AB$ are fulfilled.

A further remark may be added:

If the two independent samples had been drawn from populations having the same correlation coefficient but unequal variances (i.e. if in equations (6) and (7) $\nu/\sqrt{\alpha\beta} = \mathrm{N}/\sqrt{\mathrm{AB}} = -\rho$), then it is easy to see that again test (23) is valid with w and W replaced by $w' = w\,(\sqrt{\alpha\beta}/\sqrt{\mathrm{AB}})$ and $W' = W(\sqrt{\alpha\beta}/\sqrt{\mathrm{AB}})$ respectively.* This is analogous to the well-known property of the e^{2z} distribution.

We now turn to the case of negative ratios w:

$$w \leqq 0.$$

From equation (20) we obtain under assumption (21) for the distribution of negative values of w

$$\Phi^-(w) = \sum_{i=0}^{k} \sum_{j=0}^{K} \frac{(2k-i)!}{(k-i)!\,k!} \frac{(2K-j)!}{(K-j)!\,K!} \frac{(i+j+1)!}{i!\,j!} 2^{-2k-2K+i+j-2} \times \left[\frac{(-w)^i}{[(1+\rho)(-w)+(1-\rho)]^{i+j+2}} + \frac{(-w)^i}{[(1-\rho)(-w)+(1+\rho)]^{i+j+2}} \right] \times (1-\rho^2)^{k+K+2}. \quad \ldots\ldots(24)$$

It is easy to see that, as $\rho^2 \to 1$, $\Phi^-(w) \to 0$ uniformly in any finite interval of non-positive w-values.

Finally, we obtain by elementary transformations for any $W \geqq 0$

$$P^- = \int_{-\infty}^{-W} \Phi^-(w)\,dw = 4^{-k-K-2}(kK)^{-1} \times \sum_{q=1}^{k+1} \sum_{p=1}^{K+1} \frac{2^{p+q}}{\mathrm{B}(k, k+2-q)\,\mathrm{B}(K, K+2-p)} \times \left[(1-\rho)^{-p}(1+\rho)^{-q} I_{\left[1/\left(1+W\frac{1+\rho}{1-\rho}\right)\right]}(p,q) + (1+\rho)^{-p}(1-\rho)^{-q} I_{\left[1/\left(1+W\frac{1-\rho}{1+\rho}\right)\right]}(p,q) \right] \times (1-\rho^2)^{k+K+2}, \quad \ldots\ldots(25)$$

which is equivalent to formula (2).

If the two independent samples had been drawn from populations with different variances but equal correlation coefficients, then what has been stated about P^+ also applies to P^-.

* Thus the calculation of $P^+(\rho)$ only differs from that described previously in that a different row of Pearson's table has to be entered. Should it, however, become necessary to calculate the best estimate of ρ, it has to be remembered that now the populations (6) and (7) may have different variances.

REFERENCES

(1) PEARSON, K. *Tables of the Incomplete Beta-function* (1934). (Biometric Laboratory, Univ. Coll. London.)

(2) FISHER, R. A. *Statistical Methods for Research Workers*, 5th ed. (1934).

(3) WOODMAN, H. E., EVANS, R. E., CALLOW, E. H. and WISHART, J. "The nutrition of the bacon pig: (1) The influence of high levels of protein-intake on growth, conformation and quality of the bacon pig." *J. Agric. Sci.* XXVI (1936), pp. 546–619.

(4) WISHART, J. and BARTLETT, M. S. *Proc. Camb. Phil. Soc.* XXVIII (1932), pp. 455–9.

(5) PEARSON, K., JEFFERY, G. B. and ELDERTON, E. M. *Biometrika*, XXI (1929), pp. 164–93.

(6) PEARSON, E. S. and WILKS, S. S. *Biometrika*, XXV (1933), pp. 353–78.

A CONTRIBUTION TO THE BIOMETRIC STUDY OF THE HUMAN MANDIBLE

By FRANK H. CLEAVER, M.A.

Crewdson Benington Student in Craniometry

1. *Introduction.* It is clear to-day that the statistical study of anthropometric data with the object of investigating racial origins and relationships requires numbers of subjects, or specimens, far in excess of those considered sufficient by earlier anthropologists. The newer methods also demand greater precision in measurement, and better control and standardization of the techniques used in collecting the data. There is far more metrical material available for the cranium than for any other part of the skeleton, and the results for it are far in advance of those for any other kind of anthropometric material. The available measurements of living series, though more extensive, are unfortunately of lesser value owing to the fact that there has been no effective control or standardization of the techniques used in determining them. Owing largely to recent work presented in papers in *Biometrika*, the mandible is now the part of the skeleton, after the cranium, which has been best described metrically. The present paper provides statistical data for four new male and three corresponding female series, viz. two English, a Punjabi (male only) and an Australian. In all there are now 17 male and 9 female series measured by following the same biometric technique, though it is clear that some of these are too small to be of permanent value by themselves. The statistical treatment applied here is the same as that of the earlier papers, and particular attention is paid to a discussion of the coefficients of racial likeness. It was not to be expected that the conditions which have to be fulfilled in interpreting these criteria of resemblance would be *precisely* the same for mandibular as for cranial material. There are clear differences between the two kinds of evidence in this respect, and a general conclusion reached is that measurements of more and longer series of mandibles are still needed in order to discover how far the bone is capable of revealing racial relationships. It appears to be less effective for this purpose than the cranium.

2. *Description of the material.* Original data relating to four series of mandibles are presented in this paper. The material consists of two London series (from Spitalfields and Farringdon Street), a Punjabi and a native Australian. The first two of these are preserved at University College, London, and permission to work on them was kindly granted by the late Professor Karl Pearson; the third and fourth, preserved in the Museum of the Royal College of Surgeons of England, were loaned by the College authorities, and I should like to take this opportunity

of thanking them for their kindness, and the staff of the College for the consideration and assistance I at all times received from them.

(*a*) The Spitalfields skeletons were dug up in 1926,* and the nature of the interment in roughly circular pits, without any orderly arrangement of the bodies, pointed to a mass burial, the result of plague, massacre, or some such catastrophe. Examination of the crania showed that they are racially homogeneous, while comparisons between the crania of this and of other series (using the method of the coefficient of racial likeness) have indicated that the Spitalfields series lies closer to Pompeians, Etruscans and the population interred in the Church of St Leonard, Hythe, than it does to any other European series available. The type is far removed from those of the Neolithic, Bronze Age and Anglo-Saxon populations of England, as well as from seventeenth-century crania excavated at Whitechapel and Farringdon Street. The Spitalfields interment is therefore one of an intrusive population, and in the absence of datable artifacts, and of any other direct archaeological evidence, it has been concluded that it took place either in mediaeval or Roman times, this assumption being based mainly on an examination of the history of the Spitalfields site. The measured series is made up by 63 male and 32 female adult mandibles, only 12 of these being associated with crania. There are 195 other mandibles from Spitalfields which were not measured, either because they are immature, or else because they are too fragmentary for the purpose. Nearly 1000 crania from the site were preserved—the majority of these being incomplete—and it is estimated that about 3000 people were buried in the excavated area.

(*b*) The Farringdon Street skeletons, of which the mandibles discussed in the present paper form part, were dug up in 1924. A detailed account of the evidence for dating the bones was prepared by Professor Karl Pearson.† As a result of his examination it is safe to say that the interment of the Farringdon Street skeletons took place during the period 1610–1722 in the graveyard of the Parish Church of St Bride, but that the majority of the interments were made between 1610 and 1666 and were mainly the results of deaths from the Great Plague, 1665. Miss Beatrix G. E. Hooke measured over 350 of the Farringdon Street crania and 67 of the mandibles. The measurements of the unsexed mandibles were published in her paper, "A Third Study of the English Skull with special reference to the Farringdon Street Crania".‡ She states that several hundred mandibles were dug up, none being attached to skulls, but that the incomplete condition of these bones, due to breakages, prevented the taking of a fairly complete set of measurements except on 67. The present writer re-examined the collection of Farringdon Street mandibles and, bearing in mind certain requirements necessary for purposes of sexing (a more detailed account of which will be found in another section

* G. M. Morant and M. F. Hoadley, "A Study of the Recently Excavated Spitalfields Crania", *Biometrika*, XXIII (1931), pp. 191–248.

† *Biometrika*, XVIII (1926), pp. 1–15.

‡ *Ibid.* pp. 1–55.

of this paper), he was able to pick out and measure 90, i.e. 23 more than Miss Hooke measured. This difference can possibly be ascribed to the fact that Miss Hooke chose for measurement only those mandibles on which all, or nearly all, the 35 measurements used at that time could be taken, while the present writer used only 16 of those measurements selected as being the most reliable (see section 3 below).

(*c*) The Australian mandibles dealt with in this paper are those in the Museum collections of the Royal College of Surgeons. These specimens were obtained from several sources at different times, and there are no series of any length among them from single burial-grounds. They are divided in the Museum catalogue according to a scheme of classification based on the modern State territories, but for purposes of statistical treatment they have been pooled, with the exception of five from the Northern Territory. The justification for this procedure is that a statistical examination of about 300 Australian crania collected from all over the continent suggests that only two racial divisions can be recognized—that from the Northern Territory, where immigration is most likely to have affected the type, and that which is spread over the enormous area of the rest of Australia.* There is, as would be expected, a close connection between these two groups. An examination of the cranial facial skeleton of the two racial groups distinguished above revealed no significant differences between them, and hence it would have been of considerable interest to compare the Northern Territory mandibles with those from the remaining area. Unfortunately, no adequate numbers were available for the former group which was demarcated from the remainder in accordance with the cranial evidence mentioned above. Of this remainder the males were first divided into two sets. The first comprises those from Western Australia (3), South Australia (17) and Victoria (16), and the second those from Queensland (14) and New South Wales (9). This was a purely arbitrary division based on geographical position, and a comparison of the two groups revealed no significant differences at all between them.† Pooling is hence justified as far as can be seen from this evidence. Nine male mandibles from unknown localities were included in the pooled series. There are 36 female adult Australian mandibles: 2 from Western Australia, 12 from South Australia, 4 from Victoria, 9 from Queensland, 6 from New South Wales and 3 from unknown localities.

(*d*) The series described as Punjabi in this paper comprises those mandibles catalogued as such at the Royal College of Surgeons, and it is made up almost entirely of male mandibles from the collection which Sir Havelock Charles

* G. M. Morant, "A Study of the Australian and Tasmanian Skulls, based on previously published Measurements", *Biometrika*, XIX (1927), pp. 417–40.

† The coefficient of racial likeness based on cranial measurements for these two groups is $1{\cdot}58 \pm {\cdot}21$ for the male and $\cdot 69 \pm {\cdot}21$ for the female series (Morant, *loc. cit.* p. 424). No great reliance can be placed on any generalizations concerning the racial composition of the whole of the Australian continent in view of the scanty nature of the material available.

presented to the Museum of the College. They belonged for the most part to inmates of the British Hospital at Lahore, where Sir Havelock Charles was a surgeon, and as such they cannot really be considered as a random sample of the population of the Punjab. The few other mandibles included in the series are said to have come from various parts of the Punjab, and the specimens of the whole collection are variously catalogued as Sikh, Jat, Pathan, etc., etc. The main scheme of classification, however, distinguishes two groups—Hindu and Mohammedan—the basis of distinction being thus religious and not ethnological. In the Punjab, besides Sikhs and Pathan immigrants from across the frontier—both Mohammedan conquering stocks—there are the Mohammedan converts. The religion of Islam seems to have taken a firm hold on the native population, and, judging from census returns, large numbers of Jats, Rajputs and Gujars were of the Mohammedan faith. Just as it is confidently asserted that in Bengal the Mohammedans are of the same racial type as the lowest castes of Hindus, so in the Punjab the former are not clearly distinguished from the Hindus, though the religious divisions appear to be of significance from a racial point of view as will be shown below. The British Hospital at Lahore would no doubt have admitted the Sikhs, the Pathans, the descendants of the old Rajput rulers, the Jat peasantry and perhaps even some of the nomad Baloches, who are supposed to be of a distinctive physical type. The present sample cannot be considered a racially homogeneous one, and it must be considered from the racial standpoint only as representing an Indian type, in contrast to the European and Australian types also dealt with in this paper. Such lack of homogeneity was evidenced when the sample itself was divided into two groups—Mohammedan and Hindu. A comparison for all characters between the male groups—made up by 27 and 22 mandibles respectively—shows that differences exceeding 3·5 times their probable errors are found only for $ml\,(\Delta/(\text{p.e.}\,\Delta) = 6{\cdot}4)$ and $C' \angle\,(4{\cdot}6)$ out of the 21 characters compared. The coefficient of racial likeness between these two series was calculated for 10 characters,* giving a crude value of $1{\cdot}61 \pm {\cdot}30$ and a reduced of $7{\cdot}47 \pm 1{\cdot}40$. There is no doubt that the total Punjabi series is racially heterogeneous, but division by religion may not be the best possible, and for practical purposes it seemed advisable to use the total group, which is far from an ideal procedure. The fact that in this paper the pooled sample has been treated like a homogeneous series may be partly responsible for the unsatisfactory results (given below) which are found when racial comparisons are made between this series from the Punjab and the Farringdon Street and other series. In all 49 male adult, 9 female adult and 2 immature Punjabi mandibles were measured.

A list of all the previous series of mandibles on which measurements have been taken in accordance with the biometric technique is given by G. M. Morant.†

* A list of the characters used is given in a footnote to Table X and the Qau Egyptian standard deviations were used in calculating the coefficient.

† *Biometrika*, XXVIII (1936), pp. 92–4.

He refers to 2 European, 4 Egyptian, and 6 Asiatic series. E. S. Martin has provided similar data for another ancient Egyptian series.* Comparisons between the four series dealt with for the first time in the present paper and the earlier material are made below.†

3. *Definitions of measurements and estimates of their accuracy.* The biometric technique for measuring the human mandible was given by G. M. Morant in *Biometrika*, XIV, pp. 253–60, in 1923. Only those measurements included in a revised list,‡ and chosen chiefly because they could be taken with greater accuracy than those in the original list, have been used. These are:

w_1. Maximum breadth outside condyles avoiding excrescences on these processes. This maximum projection may be taken in any direction and it is not necessarily horizontal or transverse.

$c_y l$. Maximum projective length of the left condyle avoiding excrescences on these processes. This may be taken in any direction.

rb'. Minimum antero-posterior breadth of the left ramus at any inclination to the horizontal, but with the posterior terminal never less than 13 mm. distant from the gonion.

$m_2 p_1$. Chord between the points on the outer left alveolar margin at the middle of the second molar (or its cavity) and at the middle of the first premolar (or its cavity).

h_1. Symphyseal height from intradental to the point farthest removed from it in the symphyseal plane, this plane being determined by anatomical appreciation.

zz. Minimum chord between the anterior margins of the right and left *foramina mentalia*.

$c_r c_r$. Coronial breadth from right coronion to left coronion. If both condyles are missing, the coronia (the tips of the coronoid processes) cannot be located with sufficient accuracy to justify the measurement being taken.

The above seven are caliper measurements and all those below, except the last, are taken with the aid of a mandible board of which photographs are given in the paper describing the technique.

$M\angle$. Mandibular angle, i.e. the angle between the standard horizontal and standard rameal planes.

$c_p l$. The projective length of the corpus.

rl. The projective length of the left ramus.

* "A Study of an Egyptian Series of Mandibles, with Special Reference to Mathematical Methods of Sexing," *Biometrika*, XXVIII (1936), pp. 149–78.

† Comparisons have not been made with the Anglo-Saxon series published by J. C. Brash, Doris Layard and Matthew Young (*ibid.* XXVII (1935), pp. 398–404) as the constants for it were not available at the time when the calculation for the present paper was carried out.

‡ The technique is described and full definitions of the measurements finally adopted are given in the Appendix of his 1936 paper cited.

ml. The maximum projective length of the mandible. Both condyles make contact with the vertical rameal wing of the board, and the solid set-square makes contact with the most advanced point of the chin.

$c_r h$. Projective height of the left coronoid process.

$m_2 h$. Projective height of the corpus at the middle point of the outer alveolar margin of the second left molar.

$R\angle$. Angle of condylar-coronoidal line with ramus tangent. If either the condylar or coronoid process is defective on the left side, then the mandible is positioned from the right side.

$g_o g_o$. Chord from left gonion to right gonion, found with small calipers, the mandible board being used to locate the gonia.

The last measurement is taken with a goniometer.

$C'\angle$. Mental angle, i.e. the angle between the standard horizontal plane and the line joining the infradental to the most anterior point in the standard sagittal plane of the symphysis (pogonion). The infradental is defined to be the mid-point of the common tangent to the two curves of the outer alveolar margins of the central incisors.

The question of personal equation in regard to the characters used in this paper was investigated by G. M. Morant in his paper "A Biometric Study of the Human Mandible".* The choice of characters recommended in that paper, chiefly on the ground that they had been shown to be the most reliable of all the characters originally defined in the technique, was accepted by the present writer who followed the amplified definitions given in the Appendix.

In order to obtain estimates of personal equation, one series—the Spitalfields —was measured on two occasions. The first set of readings (Cl_1) were those obtained by the writer when he was starting to measure mandibles, and the second set (Cl_2) comprise those readings taken by the writer after nearly three months' experience of the measurements in the Galton Laboratory. Part of the data in Table I is based on 50 comparisons of these two sets for each measurement. The measurements (H) taken on the Farringdon Street mandibles by Miss B. G. E. Hooke† were also available for comparison with the measurements (Cl), 16 in number, taken by the present writer on the same series. The remainder of the data in Table I is based on comparisons of these latter two sets for each measurement. Miss Hooke followed the original definitions given in *Biometrika*, XIV (1923), and, unlike the present writer, she did not work under the direct supervision of Dr G. M. Morant. Comparisons in connection with personal equation can further be made with the data for repeated measurements published by him‡. He deals with the personal equation involved in taking two series of

* *Biometrika*, XXVIII (1936).
† *Ibid.* XVIII (1926).
‡ *Loc. cit.* (1936), Table I.

measurements himself on the same material ($M_1 - M_2$), an interval of two years having elapsed between the times when the first and second measurements were taken, and that involved when his measurements are compared with those of Miss M. Collett on the same series ($M_1 - C$). The measurements of all the above series were taken on mixed series of male and female mandibles, as it is reasonable to assume that the personal equation is likely to be the same for both sexes.

TABLE I

Data for estimating the personal equation of mandibular measurements

Characters	Maximum individual differences		Differences of means (Δ)		Standard deviations of differences	
	$H-Cl$*	Cl_1-Cl_2†	$H-Cl$	Cl_1-Cl_2	$H-Cl$	Cl_1-Cl_2
w_1	+1·4	+1·1	+0·04 ± ·073 (30)	−0·08 ± ·044 (50)	0·59 ± ·051	0·46 ± ·031
$g_o g_o$	+1·7	+&−1·8	+0·36 ± ·055 (51)	+0·13 ± ·073 (50)	0·58 ± ·039	0·77 ± ·052
$c_r c_r$	−1·3	—	+0·01 ± ·051 (36)	—	0·45 ± ·036	—
zz	+1·6	+&−0·9	−0·15 ± ·043 (50)	+0·07 ± ·030 (50)	0·45 ± ·030	0·31 ± ·021
$c_y l$	−1·1	+1·5	−0·21 ± ·045 (32)	−0·13 ± ·034 (50)	0·38 ± ·032	0·36 ± ·024
ml	+4·1	−1·5	+0·80 ± ·092 (47)	+0·03 ± ·063 (50)	0·94 ± ·065	0·66 ± ·045
$c_p l$	+3·3	+1·2	+1·04 ± ·094 (51)	+0·08 ± ·041 (50)	1·00 ± ·067	0·43 ± ·029
rb'	+2·0	−0·9	+0·29 ± ·054 (47)	+0·10 ± ·022 (50)	0·55 ± ·038	0·23 ± ·016
$m_2 p_1$	−2·2	+&−0·8	−0·07 ± ·095 (33)	−0·12 ± ·032 (50)	0·81 ± ·067	0·34 ± ·023
h_1	−3·3	—	−0·44 ± ·088 (29)	—	0·70 ± ·062	—
$m_2 h$	−1·2	+2·3	−0·27 ± ·075 (23)	−0·02 ± ·059 (50)	0·53 ± ·053	0·62 ± ·042
$c_y h$	−1·5	−1·7	−0·26 ± ·053 (46)	−0·06 ± ·054 (50)	0·53 ± ·037	0·57 ± ·038
rl	+2·7	+1·7	+0·75 ± ·077 (47)	+0·02 ± ·069 (50)	0·78 ± ·054	0·72 ± ·049
$M\angle$	+3°·5	−3°·0	−0°·16 ± ·110 (51)	−0°·60 ± ·081 (50)	1°·16 ± ·077	0°·85 ± ·057
$R\angle$	−3°·0	−2°·0	+0°·53 ± ·131 (39)	−0°·40 ± ·083 (50)	1°·21 ± ·092	0°·87 ± ·059
$C'\angle$	+6°·0	—	+1°·20 ± ·385 (28)	—	3°·02 ± ·272	—

* Differences for the Farringdon Street series. † Differences for the Spitalfields series.

The maximum differences found are given in columns 2 and 3 of Table I of the present paper, and in the same columns of Table I in Morant's paper. These maximum differences are of much the same order for the sets of differences ($M_1 - C$), ($M_1 - M_2$) and ($Cl_1 - Cl_2$), but for the set ($H - Cl$) they are considerably larger for the characters h_1, ml, $c_p l$, $m_2 p_1$ and $C'\angle$. Restricting comparisons to the 13 characters whose differences are available for all four sets, it is found that for seven of these ($H - Cl$) has the greatest maximum difference. The effect of personal equation on mean values may now be considered. Columns 4 and 5 of Table I of the present paper and of Table I of Morant's paper give the differences of the means (Δ) for the four sets of differences and the probable errors of these constants. In the set of differences ($M_1 - C$) 9 of the total 16 characters have values which differ from zero by less than three times their probable errors. In the set ($M_1 - M_2$) there are 11 out of the 16 showing the same relationship. In the set ($Cl_1 - Cl_2$) there are 8 out of 13 showing the same relationship. In the set

$(H-Cl)$ 4 characters only from a total of 16 have values which differ from zero by less than three times their probable errors. It is again evident that the differences $(H-Cl)$ are clearly distinguished from the other three sets. It will be sufficient in comparing mean values of the differences for the different sets to use the (M_1-C) values and to ignore the values of (M_1-M_2), as it has been shown by Morant that the former are on the whole the more reliable. In the comparison of Δ_{M_1-C} and $\Delta_{Cl_1-Cl_2}$ for the 13 characters possible, 2 of the differences, irrespective of signs, exceed three times their probable errors. These are for $c_r h$ (difference/p.e. difference = 3·5) and $M\angle$ (3·8). In the first of these cases the difference (M_1-C) is the greater, but for $M\angle$ the reverse is true. Thus the only character for which the standard of accuracy of the present writer is appreciably less than that previously demanded—a demand made in consideration of Dr Morant's statistical treatment of the measurements dealt with in his paper—is $M\angle$. An explanation of this fact is found in the circumstances under which the measurement of this character took place. The Spitalfields series, on which the measurements were taken to obtain the set of differences (Cl_1-Cl_2), contains a large percentage of mandibles lacking a condylar process, or having one of these processes badly damaged. The positioning of the mandible for the measuring of $M\angle$ is a matter of approximation in these cases, and such approximation on the mandible board is quite likely to give rise to unusually large errors in taking the measurements. Now, after subsequent laboratory experience, the present writer would not include measurements as doubtful as some of those recorded for the Spitalfields series, and therefore it seems reasonable to suggest that the inaccuracies consequent upon taking too many doubtful measurements are responsible for the unsatisfactory nature of his earlier readings of $M\angle$. Comparisons irrespective of signs may now be made between Δ_{M_1-C} and Δ_{H-Cl}. Out of the 16 comparisons possible 6 differences are greater than 3·5 times their probable errors. These are for $c_y l$ (3·9), $g_o g_o$ (4·3), $c_r c_r$ (5·3), ml (6·8), rl (7·6) and $c_p l$ (8·0). In all except the third of these cases Δ_{H-Cl} is greater than Δ_{M_1-C}. Three of these measurements are taken on the mandible board, and it is clear that Miss Hooke had a conception of the definitions different from that of the present writer, all her measurements tending to be greater than his.* A detailed comparison of the differences of the means (Cl_1-Cl_2) and $(H-Cl)$

* That this difference in interpretation of the definitions exists in the case of these two workers is further illustrated by an examination of the numbers of mandibles on which either took any one of the 3 significantly different mandible board measurements, while the other omitted it. Such an examination shows that Miss Hooke took 23 measurements of these characters where the present writer did not, and that the latter took 11 of the bilateral measurements (rl) on the right where Miss Hooke took them on the left. It is clear that all estimates of personal equation are likely to be considerably influenced by measurements taken on imperfect specimens, and that variabilities of differences will be very much reduced if questionable readings are omitted. In the present instance wider discrepancies would have been evident in the set of differences $(H-Cl)$, but for the fact that throughout the measurement of the whole series the present writer refrained from taking measurements in 104 doubtful cases for which Miss Hooke had given readings.

need not be made, for in 10 cases out of 13 the latter are the greater. It is thus again evident that the differences $(H - Cl)$ are of far greater account than any other differences available.

Comparing the standard deviations of the differences in the same way, it is found that for 4 of the 13 characters the differences $(M_1 - C)$ and $(Cl_1 - Cl_2)$ exceed 3·5 times their probable errors. These are for m_2h (3·7), g_og_o (3·8), rb' (5·6) and c_yl (5·9). In all cases except the third the standard deviation $(Cl_1 - Cl_2)$ is in excess. A comparison of the standard deviations of the differences $(M_1 - C)$ and $(H - Cl)$ shows that 7 are significant from a total of 16 characters. These are for c_yl (5·1), m_2p_1 (5·2), $R\angle$ (5·4), $C'\angle$ (5·5), zz (5·9), c_pl (7·3) and w_1 (7·4). In all these cases the standard deviation $(H - Cl)$ is the greater. A detailed comparison of the standard deviations of the differences for $(Cl_1 - Cl_2)$ and $(H - Cl)$ need not be made for, just as in the case of the differences of the means, in 10 comparisons out of 13 the $(H - Cl)$ constant is greater than the corresponding $(Cl_1 - Cl_2)$ constant.

It is evident, from the comparisons made between mean differences and standard deviations of differences, that the two sets of readings taken on the same English mandibles by the present writer indicate errors of personal equation which are almost precisely of the same order as those found between the readings taken by Miss Collett and G. M. Morant on an Egyptian series. Where the most significant differences between the corresponding constants were found —viz. in the case of a few of the standard deviations of the differences—the readings $(Cl_1 - Cl_2)$ are slightly less consistent than the readings $(M_1 - C)$. It was shown by Morant in his paper that the errors indicated in the latter case $(M_1 - C)$ for the characters used in the present paper are not large enough to invalidate inter-racial comparisons, and it seems safe to assume that the same will be true for the readings taken by the present writer.*

* The measurements used in this paper were selected by Morant from a larger number originally defined mainly on the grounds that they were found to be the most accurate ones. The tests used in making the selection depended on comparison of the differences of means found between two sets of readings on the same mandibles with the probable errors of the means of an Egyptian series, and on a second comparison of the standard deviations of the differences with the standard deviations of the same Egyptian series. Full details of the method used are given in his paper. Applying the same tests to the set of differences $(Cl_1 - Cl_2)$, it is found that 4 out of 13 characters fall short of the standard accepted in that paper. These are c_yl, m_2h, $M\angle$ and m_2p_1. The last did not satisfy the tests in the case of Morant's own data, but its continued use was recommended, since it is but little less reliable than the other characters accepted and it is a measurement of particular interest. The lack of reliability in the case of the characters c_yl and m_2h may perhaps be explained by the failure on the part of the present writer to reject the specimens on which it was doubtful whether a close enough approximation of the measurement could be obtained, and by his inability to deal effectively with the condylar anomalies met with in taking the measurement c_yl on the Spitalfields mandibles—the first series he measured. In the case of the differences $(H - Cl)$ 10 of the 16 characters fail to fulfil the requirements of the tests: and for this comparison the measurement $C'\angle$ is found to be the least reliable. It is, therefore, of interest to note that for the differences between Dr Morant's and the author's readings for this angle the tests are satisfied, although they were not satisfied for Morant's and Collett's original data (see Table II below).

When my readings on the Farringdon Street mandibles are compared with those of Miss Hooke on the same series, differences of a markedly higher order are found, whether the means or the standard deviations of the differences are considered. The same discordance is found when the set of differences $(H - Cl)$ is compared with the set $(M_1 - C)$. It is, therefore, reasonable to suppose that Miss Hooke's measurements on any series, when compared with those of any one of the other three measurers on the same series, would not have agreed with theirs as closely as theirs would have with one another. It is not possible to assert that she measured less *accurately* than they did, since the relations observed above between her measurements and those of the other three measurers are probably due to the fact that she had only the original unrevised definitions as a guide to her measuring, and that she was not able to work in consultation with anyone who had previously applied the technique.

To close this section on personal equation, a comparison is made between the measurement of $C' \angle$ (hitherto regarded as the most unsatisfactory of the characters included in the technique used in the present paper), taken by Morant and by the writer. This measurement was recorded for three series of mandibles by both these workers on account of its suspected unreliability. The results are set out in Table II below. The mean difference for the combined series differs significantly from zero, but both it and the standard deviation of the differences are less than the corresponding constants $(M_1 - C)$, though not significantly so.

TABLE II

Data for estimating the personal equation of the mental angle $(C' \angle)$

Series	Maximum individual differences	Differences of means (Δ)	Standard deviations of differences
	$M - Cl$	$M - Cl$	$M - Cl$
Australian	$+ \& - 2°{\cdot}5$	$+0°{\cdot}51 \pm {\cdot}096$ (62)	$1°{\cdot}12 \pm {\cdot}068$
Punjabi	$+2°{\cdot}5$	$+0°{\cdot}41 \pm {\cdot}138$ (28)	$1°{\cdot}08 \pm {\cdot}097$
Farringdon Street	$+3°{\cdot}0$	$+0°{\cdot}29 \pm {\cdot}114$ (43)	$1°{\cdot}11 \pm {\cdot}081$
Combined	$+3°{\cdot}0$	$+0°{\cdot}42 \pm {\cdot}065$ (133)	$1°{\cdot}11 \pm {\cdot}046$

On comparison with the constants of the set of differences $(H - Cl)$ in Table I, for the character $C' \angle$, it is found that the mean for the combined differences $(M - Cl)$ is the smaller, though not significantly so, while the standard deviation of the differences for $(M - Cl)$ is also the smaller, and markedly so, the difference for $(H - Cl)$ and $(M - Cl)$ in this case being 6·9 times its probable error. From consideration of the above results, it seems reasonable to assume that a marked improvement has occurred in the accuracy with which the mental angle has been

taken. Apparently the difficulties attending the measurement of this important character can be overcome by care and practice in measurement, and further improvement in accuracy would be expected if better designed instruments were substituted for those now in use, which are far from satisfactory.

4. *Methods of sexing the material.* In cases where the sexes are not known, it is probably impossible to secure absolute accuracy in sexing a series of crania or mandibles by anatomical inspection or any other method. Assessment of the sex depends on the nature of the whole series dealt with, and not on standardized conceptions of characters of the bone that remain constant for every possible series. The sexing of any series of bones available is, however, of great importance, as the statistical treatment of osteometric material demands that each sex be considered separately. Mathematical methods of sexing have therefore been devised to supplement anatomical sexing, such methods being based on the combined values of certain metrical characters of the bones. It can be assumed that the distribution of any of these particular characters for either sex, in the case of a homogeneous series, will be approximately normal. The most suitable characters for discriminatory purposes are those whose means differ most in proportion to the standard deviations of the distributions for the two sexes. The characters chosen should also have low intra-racial correlations, and it is an advantage if they can be found for a high percentage of bones, so that as few specimens as possible will have to be left unsexed.

Dr E. S. Martin has discussed several methods of sexing mandibles in a recently published paper.* He came to the conclusion that the most effective characters for sexing purposes, and those which fulfilled the above conditions most adequately, were $g_o g_o$, $c_p l$, $c_r h$ and $M\angle$, and these have been used for the purpose in the present paper. He also showed that anatomical sexing is far more reliable than had been generally supposed, the percentage agreement between mathematical and anatomical sexing being so high that considerable reliance can be placed on anatomical sexing alone. The method finally used in sexing two of the series of mandibles dealt with in this paper was a combination of a mathematical method with that of anatomical inspection in the cases where the sex was doubtful. Dr Martin demonstrated that very little difference is made to the accuracy of mathematical sexing by the inclusion of aged mandibles in a series, and hence no account has been taken of the relative ages of the adult mandibles. The sexing of the mandibles dealt with in this paper was carried out before Dr Martin's work was published, and hence the more elaborate methods of sexing given in his paper were not applied. Moreover, the series dealt with here are so short that it seems reasonable to suggest that there would be very little practical advantage in applying the more elaborate methods which he found to give only slightly better results. It will, however, be interesting to note how far the

* *Loc. cit.*

admittedly cruder mathematical methods here employed give results in agreement with anatomical sexing.

For the purpose of sexing the mandibles, it was assumed that the proportion of males to females in the sample to be sexed is the same as that of the cranial series to which it belongs. The less crude of the two mathematical methods of sexing used—referred to here as method I—is concerned, theoretically, with the sum of the four ratios obtained by dividing the deviations from the means of the characters, $g_o g_o$, $c_p l$, $c_r h$ and $M\angle$, by the corresponding standard deviations for the total series. The mean values for the characters $g_o g_o$, $c_p l$ and $c_r h$ are higher for males than females, but the reverse is true for $M\angle$, and hence the ratio for this last character has to be subtracted from the sum of the ratios of the other three for each mandible. There are 95 Spitalfields mandibles, and if the proportions of the sexes are to be supposed the same for these as for the crania we must take 32 as female and 63 as male.* The 32 with the lowest scores will be counted female. In actual practice, however, the absolute measurements themselves, and not their deviations from the means, were divided by the standard deviations for the total series, since the mandibles are arranged in the same order by these two procedures and the former entails less calculation.

The second method tried proceeds as follows: the measurements of each mandible for the characters $g_o g_o$, $c_p l$, $c_r h$ were added together, and in each instance the measurement of $M\angle$ was subtracted from this total despite the difference of units of measurement used, viz. millimetres for the first three and degrees for the last character. The 32 mandibles with the lowest totals were classed as female. This method will be referred to as method II.

The Spitalfields series was sexed anatomically by Dr G. M. Morant, who in doing this accepted the proportions of males to females given by the crania, and it will be interesting to note the percentage agreements between the methods used. These are 87·4 between inspection and method I; 85·3 between inspection and method II; and 95·8 between method I and method II. The present writer also sexed the Spitalfields series anatomically, and there is an agreement between this sexing and that obtained from the application of method I in 83·2 per cent. of cases. It is surprising to find that such a high percentage agreement between anatomical and mathematical sexing is obtained in the case of method I. This is a crude method, since the four characters used are assumed to be of equal importance for sexing purposes and no account is taken of correlations between them. It is more surprising still to find a percentage agreement nearly as high when method II is used, since this method is wholly unsatisfactory from a theoretical point of view.† As the above percentage agreements between anatomical and

* Of the 883 adult Spitalfields crania which were sexed, 590 (66·8 per cent.) were supposed male and 293 (33·2 per cent.) female.

† If either $c_r h$ or $c_p l$ is taken singly as a criterion of sex there is agreement between these estimates based on single characters and anatomical sexing of 76·8 per cent. in both cases, and an agreement between the two of 66·3 per cent.

metrical estimates are of the same order as those obtained from the more elaborate methods discussed by Dr Martin, it seems fairly reasonable to conclude that the crude mathematical method (I) is satisfactory. The mandibles of the Spitalfields series which had been classed oppositely by method I and that of anatomical inspection were re-examined anatomically, and on this re-examination the sex of the doubtful cases was decided. The crania to which 12 of the mandibles of the Spitalfields series had been attached were known, and these had been sexed anatomically. The accepted sexes of these 12 mandibles (all those numbered under 1000) were now compared with the results obtained from an anatomical sexing of the corresponding skulls, and in 9 cases agreement was found. Three cases disagreed, viz. No. 509 sexed as male mandible—skull sexed female; No. 401 sexed female mandible—skull sexed male; No. 507 sexed female mandible —skull sexed male. In each of these cases both methods of sexing the mandible gave the same result, which was accepted in spite of the disagreement with the sexing of the cranium.

The original population from which the Farringdon Street series, comprising 90 mandibles, was drawn is represented by 381 crania sexed in the ratio 213 female and 168 male. Using the same ratio we thus obtain for the mandibles 50 female and 40 male. The percentage agreement obtained between the mathematical method (I) of sexing described and that of anatomical inspection by Dr G. M. Morant was 88·9 per cent., and by the writer 86·7. There was an agreement of 86·7 per cent. between the two anatomical estimates. A re-examination of the 10 doubtful cases anatomically, as in the case of the Spitalfields series, finally decided their sexes.

A percentage agreement between anatomical and mathematical methods of sexing lies at best somewhere between 85 and 90 per cent., and it is hardly possible to improve upon this on account of the presence of border-line cases in the samples.* The most satisfactory way to sex a series of mandibles, for which the ratio of males to females is assumed known, seems to be that of re-examination of those mandibles which have been sexed oppositely by the two methods, and finally sexing these cases mainly on anatomical grounds. The mathematical method is then merely a subsidiary one used to support anatomical sexing. It should be noted that the method of sexing adopted in the case of the two English series depends essentially on the assumption that the proportions of the sexes are the same for the mandibular as for the corresponding cranial samples. If this assumption is incorrect some of the bones will inevitably be sexed incorrectly.

An examination of the sex ratios of the absolute measurements (i.e. male mean/female mean) for the Spitalfields, Farringdon Street and Australian series shows that for any particular character there is a close agreement between these

* The Australian series of mandibles—of which the sexes were known from the cranium, from the skeleton, or from more direct evidence—were sexed anatomically by Dr G. M. Morant and the writer. The anatomical sexing agreed with the sex assigned in the catalogue in 88 per cent. of cases in Dr Morant's assessment, and in 86 per cent. of cases in my own.

ratios for the different series. The small numbers that go to make up most of the female series, however, give very little weight to any conclusions drawn from such ratios. Nevertheless, it is interesting to note that the generally accepted idea that the differences of sex are more marked among primitive than among civilized peoples is not borne out by the ratios for the three series concerned, the Australian ratio being the greatest only in 4 (and equal to the Farringdon Street in 2) out of 13 comparisons. The heights have the largest ratios, but they are not markedly higher than those for the remaining measurements. The sex ratios tend to be higher for mandibular than for cranial characters in the case of a particular series. This has been shown in the case of the Kerma from Egypt, and it is also true for the Spitalfields and Farringdon Street series.

5. *Racial differences in variability.* It is possible to draw an unambiguous conclusion as to relative racial variability with respect to a measurable character of one series compared with the corresponding character of another, and one approach to the problem of variability would be to make no statement concerning relative racial variability, unless it referred merely to variabilities of single characters in different series. An alternative method of approach is to assume that an estimate of general variability for all characters measured can be obtained. When, however, a large number of characters are considered, any statement about relative total variability (i.e. variability for all characters considered together) must be somewhat arbitrary, since it depends on the particular method of comparison employed. When one race shows a consistently higher variability than another, in the case of all characters showing a significant difference, we can reasonably assume that it is the more variable. When an equal number of significantly different characters are found for each race in excess of the other, then it seems impossible to assign greater variability to either.

In Tables III and IV are given the standard deviations for all characters and the coefficients of variation for the absolute measurements, respectively. Using these coefficients for absolute measurements and the standard deviations for angles and indices, a comparison for characters considered singly may be made between the four series of mandibles for which constants are given in the tables. The results of such comparisons for the four new series and based on the 21 characters are set out in Table V. In column 2 is given, for each series in a particular comparison, the number of characters having the greater constants of variability. In column 3 are shown the characters which differ significantly in the comparison, and, where such a significant difference occurs, the series for which the variability of the significant character is the greater. An examination of Table V shows that it would be unsafe to infer from the results for the male series that one race is definitely more variable than another, since few markedly significant differences are found. An examination of the comparisons for the female series does incline one to assume that the Spitalfields series is less variable than

TABLE III

*Standard deviations for series of mandibles**

Character	Standard deviations						
	Male				Female		
	Spitalfields	Farringdon Street	Punjabi	Australian	Spitalfields	Farringdon Street	Australian
w_1 —	5·40 ± ·49	3·75 ± ·37	5·77 ± ·40	5·97 ± ·39	3·77 ± ·42	5·16 ± ·45	6·28 ± ·59
$g_o g_o$ —	6·87 ± ·41	6·58 ± ·50	5·21 ± ·36	7·83 ± ·49	5·14 ± ·43	5·74 ± ·39	5·61 ± ·56
$c_r c_r$	5·38 ± ·38	5·44 ± ·48	4·56 ± ·32	5·97 ± ·39	3·56 ± ·35	4·75 ± ·38	6·30 ± ·57
zz	2·43 ± ·15	2·21 ± ·17	2·19 ± ·15	2·49 ± ·15	2·24 ± ·19	2·56 ± ·17	2·79 ± ·22
$c_y l$	1·68 ± ·14	1·72 ± ·14	1·83 ± ·12	2·01 ± ·12	1·32 ± ·13	1·55 ± ·11	2·29 ± ·20
ml —	5·41 ± ·40	5·51 ± ·45	6·27 ± ·44	3·92 ± ·24	4·50 ± ·41	5·97 ± ·43	4·88 ± ·43
$c_p l$ —	3·98 ± ·24	3·76 ± ·28	4·54 ± ·32	5·05 ± ·31	3·10 ± ·26	3·93 ± ·27	4·55 ± ·39
rb'	2·37 ± ·14	2·60 ± ·20	2·59 ± ·18	3·01 ± ·19	1·98 ± ·17	2·61 ± ·18	2·79 ± ·23
$m_2 p_1$	1·94 ± ·13	1·22 ± ·12	1·63 ± ·13	1·52 ± ·09	1·06 ± ·12	1·44 ± ·16	1·42 ± ·12
h_1	2·49 ± ·18	2·34 ± ·32	2·15 ± ·24	2·97 ± ·20	3·07 ± ·32	2·73 ± ·26	2·36 ± ·23
$m_2 h$	1·68 ± ·11	2·88 ± ·32	2·72 ± ·23	2·42 ± ·15	2·29 ± ·24	2·60 ± ·36	2·43 ± ·22
$c_r h$	5·05 ± ·30	4·38 ± ·33	5·20 ± ·36	5·13 ± ·32	3·86 ± ·33	4·91 ± ·33	4·93 ± ·40
rl —	5·02 ± ·33	3·54 ± ·28	4·22 ± ·29	5·10 ± ·32	3·26 ± ·29	4·23 ± ·31	5·39 ± ·46
$M\angle$ —	7°·03 ± ·42	5°·67 ± ·43	5°·65 ± ·39	6°·55 ± ·41	5°·84 ± ·49	6°·21 ± ·42	5°·01 ± ·43
$R\angle$	7°·74 ± ·53	8°·50 ± ·67	7°·88 ± ·54	7°·10 ± ·44	6°·08 ± ·56	8°·96 ± ·62	5°·69 ± ·50
$C'\angle$	6°·67 ± ·47	7°·99 ± ·95	8°·36 ± ·85	5°·68 ± ·43	6°·73 ± ·70	6°·13 ± ·56	5°·16 ± ·55
$100\, c_r h/ml$	4·77 ± ·35	5·73 ± ·47	6·16 ± ·43	5·47 ± ·35	3·94 ± ·36	4·79 ± ·35	4·73 ± ·42
$100\, c_r c_r/ml$	6·27 ± ·51	7·67 ± ·73	6·02 ± ·43	6·69 ± ·44	4·98 ± ·50	6·03 ± ·53	7·14 ± ·67
$100\, g_o g_o/c_p l$	11·59 ± ·70	11·55 ± ·87	10·37 ± ·72	11·92 ± ·77	10·98 ± ·93	10·72 ± ·72	9·70 ± ·96
$100\, rb'/rl$	5·14 ± ·34	4·82 ± ·38	7·01 ± ·49	4·99 ± ·31	4·06 ± ·37	6·45 ± ·47	6·23 ± ·53
$100 g_o g_o/c_r c_r$	7·25 ± ·51	8·57 ± ·76	5·51 ± ·39	8·51 ± ·59	6·33 ± ·62	7·53 ± ·61	8·19 ± ·83

* The numbers of mandibles on which the constants in this table and in Table IV are based can be seen from the table of means (Table IX).

TABLE IV

Coefficients of variation for series of mandibles

Character	Coefficients of variation						
	Male				Female		
	Spitalfields	Farringdon Street	Punjabi	Australian	Spitalfields	Farringdon Street	Australian
w_1	4·50 ± ·41	3·19 ± 0·32	4·96 ± ·35	4·04 ± ·27	3·33 ± 0·37	4·73 ± 0·42	5·65 ± 0·53
$g_o g_o$	7·13 ± ·43	6·73 ± 0·51	5·61 ± ·39	8·21 ± ·52	5·86 ± 0·49	6·70 ± 0·45	6·47 ± 0·64
$c_r c_r$	5·57 ± ·39	5·67 ± 0·51	4·84 ± ·34	6·30 ± ·41	3·92 ± 0·38	5·18 ± 0·42	7·17 ± 0·65
zz	5·39 ± ·33	5·03 ± 0·38	4·99 ± ·34	5·25 ± ·30	5·16 ± 0·44	5·94 ± 0·40	6·09 ± 0·49
$c_y l$	8·19 ± ·67	8·69 ± 0·72	8·97 ± ·61	9·39 ± ·59	6·95 ± 0·68	8·61 ± 0·62	11·80 ± 1·03
ml	5·28 ± ·39	5·29 ± 0·43	6·12 ± ·43	3·63 ± ·22	4·57 ± 0·41	6·01 ± 0·44	4·77 ± 0·42
$c_p l$	5·37 ± ·32	4·93 ± 0·37	6·10 ± ·43	6·07 ± ·38	4·58 ± 0·39	5·62 ± 0·38	5·99 ± 0·52
rb'	7·36 ± ·44	8·41 ± 0·63	8·41 ± ·58	8·77 ± ·52	6·87 ± 0·58	9·22 ± 0·62	8·83 ± 0·73
$m_2 p_1$	6·93 ± ·49	4·33 ± 0·44	5·70 ± ·47	5·00 ± ·30	4·03 ± 0·44	5·16 ± 0·56	4·86 ± 0·42
h_1	7·71 ± ·55	7·57 ± 1·05	6·44 ± ·72	8·92 ± ·61	10·62 ± 1·12	9·19 ± 0·86	7·66 ± 0·74
$m_2 h$	6·40 ± ·44	11·57 ± 1·28	10·54 ± ·88	9·20 ± ·57	9·79 ± 1·03	11·02 ± 1·54	9·92 ± 0·90
$c_r h$	7·79 ± ·47	6·75 ± 0·51	7·90 ± ·56	8·02 ± ·50	6·75 ± 0·57	8·67 ± 0·58	8·82 ± 0·72
rl	8·04 ± ·53	5·69 ± 0·45	6·75 ± ·47	8·11 ± ·51	5·72 ± 0·52	7·91 ± 0·58	9·57 ± 0·83

the Farringdon Street and the Australian, since the former shows the lesser variability for most of the characters in these comparisons.

A comparison (see Table VI) was next made for another group of four series, viz. two Egyptian (the Egyptian *E* and Kerma, the latter having been stated to be slightly more variable than another Egyptian series from Qau*), and the Australian and Spitalfields, which are assumed on such evidence as is afforded in Table V to be the most and the least variable, respectively, of the series dealt with

TABLE V

Comparisons of the variabilities of two English, an Indian, and an Australian series of mandibles†

Series	Nos. of greater constants	Significant differences
Male:		
Spitalfields (SF.) : Farringdon Street (FA.)	SF. 11>, FA. 10>	m_2p_1 (3·9) SF.>, m_2h (3·8) FA.>
Spitalfields : Punjabi (Pu.)	SF. 10>, Pu. 11>	m_2h (4·2) Pu.>
Spitalfields : Australian (Aus.)	SF. 8>, Aus. 13>	ml (3·7) SF.>, m_2h (3·8) Aus.>
Farringdon Street : Punjabi	FA. 10>, Pu. 10>, 1=	w_1 (3·7) Pu.>, 100 rb'/rl (3·5) Pu.>, 100 g_og_o/c_rc_r (3·6) FA.>
Farringdon Street : Australian	FA. 7>, Aus. 14>	rl (3·6) Aus.>
Punjabi : Australian	Pu. 9>, Aus. 12>	g_og_o (4·0) Aus.>, ml (5·2) Pu.>, 100 g_og_o/c_rc_r (3·9) Aus.>
Female:		
Spitalfields : Farringdon Street	SF. 3>, FA. 18>	100 rb'/rl (3·9) FA.>
Spitalfields : Australian	SF. 5>, Aus. 16>	w_1 (3·6) Aus.>, c_rc_r (4·3) Aus.>, rl (3·9) Aus.>, c_yl (3·9) Aus.>
Farringdon Street : Australian	FA. 12>, Aus. 9>	$R\angle$ (4·1) FA.>

† The constants compared are coefficients of variation of the absolute measurements (Table IV) and standard deviations of the indices and angles (Table III).

there. There seems to be no marked difference between the Kerma and the Australian series, though the Kerma, it is interesting to note, appears to be considerably more variable than the other Egyptian series (Egyptian *E*), as does also the Australian, especially in the comparison for the female series. A comparison of the two series which appear to show the least variability (viz. the

* G. M. Morant, *loc. cit.* p. 102.

TABLE VI

Comparisons of the variabilities of two Egyptian, an Australian, and an English series of mandibles

Series	Nos. of greater constants	Significant differences
Male:		
Kerma (K.): Australian (Aus.)	K. 10>, Aus. 11>	—
Kerma : Egyptian *E* (Eg.)	K. 16>, Eg. 5>	rl (4·2) K.>
Kerma : Spitalfields (SF.)	K. 15>, SF. 6>	—
Egyptian *E*: Australian	Eg. 6>, Aus. 14>, 1=	$c_y l$ (4·6) Aus.>, ml (4·1) Eg.>
Egyptian *E*: Spitalfields	Eg. 8>, SF. 12>, 1=	$m_2 h$ (4·6) Eg.>
Spitalfields : Australian	SF. 8>, Aus. 13>	ml (3·6) SF.>, $m_2 h$ (3·8) Aus.>
Female:		
Kerma : Australian	K. 11>, Aus. 10>	$M\angle$ (3·7) K.>
Kerma : Egyptian *E*	K. 17>, Eg. 3>, 1=	100 $c_r h/ml$ (3·6) K.>, $c_r h$ (5·0) K.>
Egyptian *E*: Australian	Eg. 7>, Aus. 14>	$c_r h$ (3·7) Aus.>, $c_y l$ (3·8) Aus.>
Egyptian *E*: Spitalfields	Eg. 14>, SF. 7>	w_1 (4·3) Eg.>, $c_r c_r$ (3·7) Eg.>
Spitalfields : Australian	SF. 5>, Aus. 16>	w_1 (3·6) Aus.>, $c_r c_r$ (4·3) Aus.>, $c_y l$ (3·9) Aus.>, rl (3·9) Aus.>

Spitalfields and the Egyptian *E*) indicates that for the female series the Egyptian *E* may be considered the more variable.

Although there appear to be racial differences in variability, the material available for the mandible accords with the far more extensive material relating to cranial and living series, in showing that the absolute differences between the variabilities of different races are exceedingly small. It is interesting to observe from Tables V and VI that the Australian and Egyptian tend to be more variable than the two English series. This is an unexpected result, but it may be due to some peculiar selection of the mandibles preserved.

6. *Special topics: correlations, asymmetry and records relating to teeth.* The coefficients of correlation between the various mandibular measurements throw some light on the interdependence during growth of various parts of the mandible.* The first structural peculiarity we note from an examination of the correlations in Morant's paper is the fact that the dental arcade between the mid-points of the alveolar margins of the first premolar and second molar ($m_2 p_1$) seems to be uncorrelated with any other chords. Apparently, growth of this particular area

* This section is complementary to that on correlation in G. M. Morant's paper, "A Biometric Study of the Human Mandible", *loc. cit.* pp. 103–8.

ceases at an early age. Thus the coefficients of correlation between m_2p_1 and all the antero-posterior chords were previously found to be insignificant: in the Australian male series, however, a significant value is found for m_2p_1 and c_pl ($r = +\cdot328 \pm \cdot081$). This would certainly have been expected *a priori* as c_pl "covers" m_2p_1.

Since the correlation coefficients for like measurements (i.e. breadths with breadths, heights with heights, etc.) are most interesting when they are insignificant, and since the converse holds good for unlike measurements, it is of interest to note (for the Qau Egyptian series) that though h_1 is, as expected, correlated with m_2h, yet it is not so with rl or with any other rameal height. The measurement (h_1) is, however, fairly highly correlated with ml, and this seems to point to the fact that an antero-posterior growth of the mandible necessitates a corresponding growth in height of the corpus. It is surprising to find a low correlation coefficient between c_pl and ml, two antero-posterior chords, especially as the former is "covered" by the latter. An unexpectedly high correlation is that between $R\angle$ and $100\,c_rc_r/ml$, though what growth factor influences these two particular measurements is not evident. Another result which reveals an unexpected feature of the architecture of the mandible is the high negative correlation found between the breadth of the ramus (rb') and the mandibular angle ($M\angle$). The broad solid ramus is apparently found on the upright looking mandible, while the slender ramus accompanies the sloping type of mandible.

The few correlation coefficients computed for the Australian male series by the writer, and set out below, lead to the same conclusions as those derived from the Qau series. A comparison between corresponding values for the two series reveals no single significant difference. The following coefficients are found for the Australian bones:

h_1 and $M\angle + \cdot090 \pm \cdot102$ (43); h_1 and $ml + \cdot516 \pm \cdot075$ (44),
h_1 and $m_2h + \cdot539 \pm \cdot071$ (45); h_1 and $m_2p_1 + \cdot015 \pm \cdot097$ (48),
h_1 and $rl + \cdot291 \pm \cdot094$ (43); rb' and $M\angle - \cdot509 \pm \cdot065$ (59),
$C'\angle$ and $M\angle - \cdot299 \pm \cdot100$ (38); m_2h and $m_2p_1 - \cdot064 \pm \cdot088$ (58),
c_pl and $ml + \cdot390 \pm \cdot076$ (57); c_pl and $m_2p_1 + \cdot328 \pm \cdot081$ (55),
$R\angle$ and $100\,c_rc_r/ml + \cdot548 \pm \cdot066$ (51).

It appears that all intra-racial correlations between absolute measurements of the mandible are positive, whether significant or not, or negative and insignificant. In other words, a large mandible tends to be large in all respects, and a small one to be small in all respects, a fact for which the normal growth of the mandible as a whole is evidently responsible. The measurement m_2p_1 alone shows no tendency to conform to the general growth trend.

The asymmetry of the mandible has sometimes been taken as a well-established fact, and anatomists as famous as Le Double have even made categorical statements to the effect that the right side of the mandible is on the average always

greater than the left. An examination of the bilateral measurements—viz. m_2p_1, c_yl and rb'*—which were taken on the mandibles of the Qau and Kerma male and female series has shown that there is fairly clear evidence of a slight asymmetry in type for these Egyptian series, and that the right side of the mandible is not consistently greater than the left. In fact, the length of part of the dental arcade (m_2p_1) as well as the breadth of the ramus (rb') were larger on the left side than on the right, the length of the condyle (c_yl) alone supporting Le Double's hypothesis. It is strange to find that the broader ramus supports the smaller condyle (at least if the length of the condyle is any criterion of its size). From Table VII it will be seen that the statistical evidence warrants no assumption of asymmetry in type in the case of the Australian mandible. The means actually found are all slightly greater on the right than on the left, but the bilateral differences are quite insignificant.

TABLE VII

Constants of bilateral differences for the Australian male series

Means ($L-R$)			Standard deviations		
m_2p_1	c_yl	rb'	m_2p_1	c_yl	rb'
$-\cdot079\pm\cdot075$ (57)	$-\cdot028\pm\cdot096$ (50)	$-\cdot006\pm\cdot098$ (64)	$0\cdot84\pm\cdot053$	$1\cdot01\pm\cdot068$	$1\cdot16\pm\cdot069$

Comparing the male Australian differences with those previously given for the two Egyptian series,† no clearly significant differences are found, so there is no evidence to show that there are racial differences in asymmetry.

Table VIII shows that a high percentage of the Australian mandibles, both male and female, had never lost a single tooth during life. This is in marked contrast to the civilized English series represented by the Farringdon Street (composed of seventeenth-century Londoners) and Spitalfields series (probably a population living in England during the Romano-British period). The Indian series has about 50 per cent. of its mandibles—25 out of 49—with a complete set of teeth, and, while being far below the corresponding percentage for the Australian series (approximately 80 per cent.), such a figure is much higher than that for the English series, which for males and females combined gives a percentage of 22·5.‡ There were several cases in the English and Indian series of arthritic condyles, though not a single instance of this was found in the Australian series which contains a case of syphilis (See Plate V A) according to the catalogue of the Royal College of Surgeons. The small numbers found in category 3 of Table VIII

* These measurements are usually taken, when possible, on the left side only.

† G. M. Morant, *loc. cit.* Table II.

‡ In the case of the long Egyptian *E* series examined by Dr Martin the percentages having all teeth, including third molars, present at death are 40·7 for males and 44·3 for females.

do not in any way affect the conclusions stated above and based on the figures found in category 2. It is to be noted that the third category of the table includes some cases for which the third molars had probably never erupted; for an examination of the dental arcade often fails to determine whether a molar was lost before death, or whether it had never erupted at all. The following cases of overcrowding were noted: Spitalfields male 3, female 2; Australian male 4, female 0; Farringdon Street male 1, female 1; Punjabi male 0. It appears that the opinion expressed by several anthropologists to the effect that overcrowding of the teeth is on the increase among modern civilized peoples is not supported by these figures, which

TABLE VIII

Comparisons of the dentitions of series of mandibles

	Male				Female		
	Spital-fields	Farring-don Street	Austra-lian	Punjabi	Spital-fields	Farring-don Street	Austra-lian
1. No. for which dental arcade is complete	27	69*	72†	49‡	22	83§	36
2. All teeth including 3rd molars present at death	10	18	58	25	5	12	26
3. All teeth except one or both 3rd molars present at death	5	8	3	1	4	8	2
4. No. having lost one or more teeth in front of molars before death	5	24	4	16	6	37	7

* 40 measured, 29 not measured; these latter were sexed anatomically.
† Including Northern Territory mandibles and 1 with socket for single pair of incisors.
‡ Including 2 with sockets for three incisors only.
§ 50 measured, 33 not measured; these latter were sexed anatomically.

refer only to the more marked cases.|| Furthermore, to assert on the alleged evidence of the increased incidence of overcrowding of the teeth that there is a tendency for modern man, civilized and uncivilized, to have a dental arcade smaller in size than that of his primitive ancestors would be unjustifiable, since overcrowding is dependent on the ratio of size of teeth to size of jaw, and not on the absolute size of the dental arcade.

7. *Comparisons by the method of the coefficient of racial likeness.* One of the primary needs of physical anthropologists in dealing with problems of human evolution is a means of classifying the races of man. To meet this need Professor Karl Pearson devised the method of the coefficient of racial likeness, which is a

|| Several cases of impaction of the third molar were also noted.

generalized statistical criterion derived from pairs of racial series and based on the comparison of a number of their mean measurements for different characters. Large numbers of these coefficients computed for different groups of cranial series have been published, chiefly in *Biometrika*, and the method has also been applied to data for series of living people in several papers. In his study of the mandible (*loc. cit.* 1936) Morant gives values for all possible pairs of 12 male and 5 female series, and he says (p. 116) that: "these criteria lead to a reasonable arrangement of the types which encourages the hope that a similar comparison of more extended material would furnish valuable aid in estimating racial relationships. At the same time the fact that samples of the sizes at present available are not differentiated cannot be accepted as a test of racial identity." The coefficients given by Martin (*loc. cit.* Table V) between his series of 26th–30th Dynasty mandibles from Gizeh and the earlier material do not conflict with these conclusions. We are now able to add comparisons with the 4 male and 3 female series described in the present paper, making a total of 17 male and 9 female, and it will be seen that these make it necessary to reconsider the position and, indeed, to question whether it is possible to obtain any rational classification of races from measurements of mandibles. The mean measurements of the new series are given in Table IX.

If M_s is the mean and σ_s the standard deviation of the sth character, these being based on n_s individuals, in the case of the first series, and if $M_{s'}$, $\sigma_{s'}$ and $n_{s'}$ are the corresponding constants for the second series, then what is now called the "crude" coefficient of racial likeness is defined to be:

$$\frac{1}{m}\Sigma\left\{\frac{(M_s-M_{s'})^2}{\frac{\sigma_s^2}{n_s}+\frac{\sigma_{s'}^2}{n_{s'}}}\right\}-1\pm\cdot67449\sqrt{\frac{2}{m}},$$

where m characters are compared. The standard deviations for the shorter series are likely to be particularly unreliable, and hence it is assumed that they are equal to those for the longest homogeneous series available. It has been shown above that, though these constants for different series show a few significant differences, yet they tend to be of closely similar orders in the case of a particular character. Supposing that $\sigma_s=\sigma_{s'}$, the coefficient becomes:

$$\frac{1}{m}\Sigma\left\{\frac{(M_s-M_{s'})^2}{\sigma_s^2}\times\frac{n_s n_{s'}}{n_s+n_{s'}}\right\}-1\pm\cdot67449\sqrt{\frac{2}{m}},$$

which is written, for convenience, as:

$$\frac{1}{m}\Sigma(\alpha)-1\pm\cdot67449\sqrt{\frac{2}{m}}.$$

Following Morant, the standard deviations of the ancient Egyptian series from Qau were used.* From the crude coefficient we can obtain a generalized measure

* The Qau is the longest of the series used in his paper. That of the 26th–30th Dynasty mandibles from a cemetery at Gizeh, described by Martin, is considerably longer, but the standard deviations for it were not available when the computation for the present paper was carried out.

TABLE IX

Mean measurements of series of mandibles

Character	Male				Female		
	Spitalfields	Farringdon Street	Punjabi	Australian	Spitalfields	Farringdon Street	Australian
w_1 —	120·0 ± ·69 (28)	117·7 ± 0·53 (23)	116·4 ± 0·57 (47)	118·5 ± 0·55 (53)	113·3 ± 0·60 (18)	110·8 ± 0·64 (30)	111·2 ± 0·83 (26)
$g_o g_o$ —	96·3 ± ·58 (63)	97·7 ± 0·70 (40)	92·8 ± 0·51 (47)	95·4 ± 0·70 (57)	87·7 ± 0·61 (32)	85·7 ± 0·55 (50)	86·7 ± 0·79 (23)
$c_r c_r$	96·5 ± ·54 (46)	95·9 ± 0·68 (29)	94·2 ± 0·45 (46)	94·8 ± 0·55 (53)	90·8 ± 0·49 (24)	91·7 ± 0·54 (35)	87·8 ± 0·80 (28)
zz	45·1 ± ·21 (61)	43·9 ± 0·24 (40)	43·9 ± 0·21 (48)	47·4 ± 0·21 (67)	43·4 ± 0·27 (31)	42·9 ± 0·25 (49)	45·8 ± 0·32 (35)
$c_y l$ —	20·5 ± ·19 (35)	19·8 ± 0·20 (34)	20·4 ± 0·18 (49)	21·4 ± 0·18 (60)	19·0 ± 0·18 (24)	18·0 ± 0·16 (45)	19·4 ± 0·28 (31)
ml —	102·5 ± ·56 (42)	104·1 ± 0·64 (34)	102·5 ± 0·62 (47)	108·1 ± 0·34 (61)	98·5 ± 0·57 (28)	99·4 ± 0·61 (43)	102·4 ± 0·61 (29)
$c_p l$ —	74·0 ± ·34 (63)	74·9 ± 0·40 (40)	74·4 ± 0·45 (47)	83·2 ± 0·44 (59)	67·7 ± 0·37 (32)	69·7 ± 0·37 (50)	75·9 ± 0·55 (31)
rb'	32·2 ± ·20 (63)	30·9 ± 0·28 (40)	30·8 ± 0·25 (49)	34·3 ± 0·25 (65)	28·8 ± 0·24 (32)	28·3 ± 0·25 (50)	31·6 ± 0·32 (34)
$m_2 p_1$	28·0 ± ·19 (47)	28·2 ± 0·18 (22)	28·6 ± 0·19 (34)	20·4 ± 0·13 (63)	26·3 ± 0·16 (19)	27·7 ± 0·22 (19)	29·2 ± 0·17 (31)
h_1	32·3 ± ·25 (45)	30·9 ± 0·46 (12)	33·4 ± 0·34 (18)	33·3 ± 0·29 (49)	28·9 ± 0·45 (21)	29·7 ± 0·36 (26)	30·8 ± 0·32 (24)
$m_2 h$	26·4 ± ·16 (49)	24·9 ± 0·45 (19)	25·8 ± 0·32 (33)	26·3 ± 0·21 (61)	23·4 ± 0·34 (21)	23·6 ± 0·51 (12)	24·5 ± 0·31 (28)
$c_r h$	64·8 ± ·43 (63)	64·8 ± 0·47 (40)	65·8 ± 0·51 (47)	64·0 ± 0·45 (60)	57·2 ± 0·46 (32)	56·5 ± 0·47 (50)	55·9 ± 0·56 (35)
rl —	62·4 ± ·47 (53)	62·2 ± 0·40 (36)	62·5 ± 0·42 (47)	62·9 ± 0·45 (59)	57·0 ± 0·42 (28)	53·5 ± 0·44 (43)	56·3 ± 0·65 (31)
$M\angle$ —	121°·5 ± ·60 (63)	121°·7 ± 0·60 (40)	120°·1 ± 0·56 (47)	117°·0 ± 0·58 (59)	126°·5 ± 0·70 (32)	127°·8 ± 0·59 (50)	121°·9 ± 0·61 (31)
$R\angle$	71°·6 ± ·75 (48)	72°·0 ± 0·94 (37)	73°·5 ± 0·77 (48)	70°·9 ± 0·63 (58)	65°·8 ± 0·79 (27)	71°·2 ± 0·87 (48)	65°·2 ± 0·71 (29)
$C'\angle$	70°·5 ± ·66 (46)	61°·8 ± 1·35 (16)	71°·2 ± 1·20 (22)	78°·0 ± 0·61 (40)	66°·4 ± 0·99 (21)	67°·3 ± 0·80 (27)	81°·6 ± 0·78 (20)
$100\ c_r h/ml$	63·3 ± ·50 (42)	62·4 ± 0·66 (34)	64·4 ± 0·61 (47)	59·3 ± 0·49 (57)	58·0 ± 0·50 (28)	56·8 ± 0·49 (43)	54·4 ± 0·59 (29)
$100\ c_r c_r/ml$	94·8 ± ·71 (35)	92·3 ± 1·03 (25)	92·0 ± 0·61 (45)	87·8 ± 0·63 (52)	92·1 ± 0·70 (23)	92·4 ± 0·74 (30)	85·8 ± 0·94 (26)
$100\ g_o g_o/c_p l$	130·6 ± ·98 (63)	130·9 ± 1·23 (40)	125·2 ± 1·02 (47)	114·9 ± 1·08 (55)	129·9 ± 1·31 (32)	123·3 ± 1·02 (50)	115·4 ± 1·36 (23)
$100\ rb'/rl$	51·8 ± ·48 (53)	50·0 ± 0·54 (36)	49·3 ± 0·69 (47)	54·7 ± 0·44 (59)	50·8 ± 0·52 (28)	53·2 ± 0·66 (43)	56·5 ± 0·75 (31)
$100\ g_o g_o/c_r c_r$	100·7 ± ·72 (46)	102·9 ± 1·07 (29)	98·8 ± 0·55 (45)	100·4 ± 0·83 (48)	96·6 ± 0·87 (24)	94·0 ± 0·86 (35)	98·8 ± 1·18 (22)

of the probability that the two samples compared were drawn from populations with identical means. The formula is theoretically more correct if the m characters used are uncorrelated with one another intra-racially. Ten were selected by Morant, principally because the correlations between them are low in the case of the male Qau series: of the total 45 r's 40 are less than ·3, the highest value is +·597 and no single character has more than two coefficients greater than ·3. The same 10 characters were used in calculating all the coefficients of racial likeness for mandibular measurements, including those given in the present paper. It may be noted that this number is considerably smaller than the 31 used, when possible, in the comparisons of cranial series by the same method.

Having the same nature as a measure of probability, the crude coefficient of racial likeness depends on the sizes of the samples compared. But the anthropologist is more interested in a measure of the absolute divergence of the types, and this is supposed to be obtained from the crude coefficients by adjusting them to values they might be expected to have if the samples were made up, not by the numbers actually available, but by 100 individuals each. If $\overline{n}_1$ and $\overline{n}_2$ are the mean numbers of individuals available for the m characters in the case of the first and second series in the comparison, respectively, then the reduced coefficient is defined to be:

$$50 \times \frac{\overline{n}_1 + \overline{n}_2}{\overline{n}_1 \overline{n}_2} \left\{ \frac{1}{m} \Sigma(\alpha) - 1 \right\} \pm 50 \times \frac{\overline{n}_1 + \overline{n}_2}{\overline{n}_1 \overline{n}_2} \times \cdot 67449 \sqrt{\frac{2}{m}}.$$

In a general way the classifications of groups of cranial series based on reduced coefficients of racial likeness that have hitherto been given accord with evidence of other kinds. The failure of the same method to give as reasonable results when applied to series of mandibles may possibly be due in this case to the inadequacy of one or other of the assumptions made in calculating the reduced coefficients. This possibility will be examined after presenting the results.

The reduced coefficients for the four new series, and between them and all the earlier ones, are given in Table X, and the values for all other pairs of the 17 male and 9 female series will be found in the papers by Morant and Martin. It has been shown repeatedly for cranial data that in attempting to derive a classification of the racial types from such material the most reasonable and suggestive results are always obtained if only the lowest orders of reduced coefficients are considered, while no account is taken of any greater than an arbitrarily defined limit. It appears to be an advantage to choose this limit as low as any which can be conveniently used for a particular group of series. In the case of the mandibular data the larger values of the reduced coefficients fail entirely to provide any arrangement of the types which could be supposed to indicate their inter-relationships and, accordingly, only the lower values will be considered now. In discussing the material available to him, Morant ignored all greater than 11, and for the material available now it was found that a limit of 10 could be used more

TABLE X

Reduced coefficients of racial likeness for mandibular series

Series	Sex	$\bar{n}$*	Spitalfields	Farringdon Street	Punjabi	Australian
Anglo-Saxon	♂	42·4	6·54 ± 0·67	17·88 ± 0·83	20·14 ± 0·70	25·40 ± 0·61
	♀	41·2	21·54 ± 0·95	33·77 ± 0·74	—	23·16 ± 0·88
Dunstable	♂	37·3	6·57 ± 0·72	11·82 ± 0·88	15·81 ± 0·75	26·26 ± 0·66
Spitalfields	♂	48·5	—	2·82 ± 0·79	5·46 ± 0·66	38·10 ± 0·56
	♀	26·0	—	16·75 ± 0·95	—	51·07 ± 1·09
Farringdon Street	♂	31·8	2·82 ± 0·79	—	2·15 ± 0·82†	49·56 ± 0·73
	♀	40·3	16·75 ± 0·95	—	—	41·99 ± 0·89
Badari Egyptian (Predynastic)	♂	33·5	40·66 ± 0·76	43·27 ± 0·92	31·04 ± 0·80	56·87 ± 0·70
	♀	18·9	53·29 ± 1·38	31·84 ± 1·17	—	42·39 ± 1·31
Qau Egyptian (4th–11th Dynasty)	♂	66·4	11·81 ± 0·54	11·09 ± 0·70	4·50 ± 0·57	32·43 ± 0·48
	♀	55·7	31·69 ± 0·85	17·18 ± 0·64	—	19·27 ± 0·79
Sedment Egyptian (9th Dynasty)	♂	32·4	16·51 ± 0·78	21·43 ± 0·94	10·15 ± 0·81	31·80 ± 0·72
	♀	21·2	29·06 ± 1·29	17·92 ± 1·08	—	40·91 ± 1·23
Kerma Egyptian (12th–13th Dynasty)	♂	55·7	24·65 ± 0·58	29·34 ± 0·74	18·28 ± 0·62	19·75 ± 0·52
	♀	44·7	50·39 ± 0·92	29·36 ± 0·71	—	29·78 ± 0·85
Gizeh Egyptian *E* (26th–30th Dynasty)	♂	211·7	5·02 ± 0·38	9·51 ± 0·55	7·48 ± 0·42	32·60 ± 0·33
	♀	131·8	10·16 ± 0·69	6·08 ± 0·48	—	39·67 ± 0·62
Tamil	♂	33·0	8·29 ± 0·77	4·94 ± 0·93	7·56 ± 0·80	34·71 ± 0·71
Punjabi	♂	43·4	5·46 ± 0·66	2·15 ± 0·82†	—	39·38 ± 0·60
Nepalese	♂	18·9	13·26 ± 1·11	10·67 ± 1·27	10·91 ± 1·15	26·77 ± 1·05
Tibetan *A*	♂	24·9	12·41 ± 0·92	8·72 ± 1·08	17·01 ± 0·96	30·25 ± 0·86
Tibetan *B*	♂	11·9	26·17 ± 1·58	36·08 ± 1·74	32·49 ± 1·61	9·71 ± 1·52
Hylam Chinese	♂	38·8	5·24 ± 0·70	14·33 ± 0·86	17·89 ± 0·74	26·39 ± 0·64
Fukien Chinese	♂	37·5	18·22 ± 0·71	30·09 ± 0·88	27·15 ± 0·75	22·16 ± 0·66
Australian	♂	59·0	38·10 ± 0·56	49·56 ± 0·73	39·38 ± 0·60	—
	♀	29·3	51·07 ± 1·09	41·99 ± 0·89	—	—

* The $\bar{n}$'s are the mean numbers of mandibles available for the 10 characters (w_1, zz, $c_y l$, ml, $m_2 p_1$, rb', h_1, rl, $R\angle$ and $100\, g_o g_o / c_y l$) used in computing the coefficients.

† The crude coefficient corresponding to this is 0·79 ± ·30.

conveniently. The arrangement suggested by the reduced coefficients can be appreciated most easily from Fig. 1, which shows all the connexions between the male series given by values less than 10. The Badari Predynastic Egyptian is the only series which has no reduced coefficient less than the arbitrary limit chosen.

The 17 series can be divided into three groups—an English, an ancient Egyptian and an Asiatic—and the Australian series. Considering these in turn, an unexpected relation is at once found in the insignificant coefficient between the Anglo-Saxon and Dunstable series, indicating that no distinction can be made

between their mandibular types although the cranial types are clearly differentiated (reduced coefficient = 13·66 ± ·45*). The mandibular value (− 0·18 ± ·76) is here less than the cranial and of an entirely different order. Equally unexpected is the linking of both the Anglo-Saxon and the Farringdon Street series to the Spitalfields, and also the lack of any connexion between the Anglo-Saxon and Farringdon Street series. The cranial evidence shows the inverse relation to this, viz. a close connexion between the Anglo-Saxons and seventeenth-century Londoners and a clear distinction between these two and the Spitalfields population, which is of uncertain date. Between the Anglo-Saxon and Farringdon Street series the cranial reduced coefficient is 8·79 ± ·32 and the mandibular 17·88 ± ·83; here the mandibular value is greater than the cranial and of a different order.

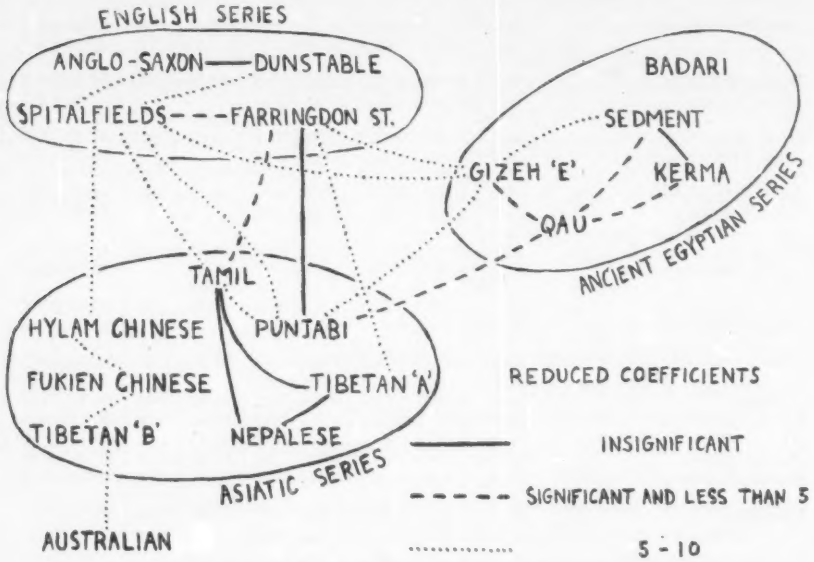


Fig. 1. The lowest reduced coefficients of racial likeness for 17 series of male mandibles.

Turning to the ancient Egyptian group, the fact that the Badari shows no connexion with any other series is not surprising, as it is the only one available of predynastic date and it is assigned to one of the earliest known predynastic periods. The insignificant coefficient between the Sedment and Kerma series (1·90 ± ·74) is unexpected, as the reduced coefficient for the cranial series is 16·41 ± ·31. For this Egyptian group, however, there are no results as unsatisfactory as those noted for the English group. The Asiatic mandibular series again show entirely unexpected resemblances and divergences. The Tamil, Nepalese

* The reduced coefficients of racial likeness for cranial series given here are all taken from papers in *Biometrika* published in 1926 or later.

and Tibetan *A* series have coefficients with one another which all differ insignificantly from zero, while the three corresponding cranial reduced values range from 12·6 to 40·9. For the Hylam and Fukien Chinese the reduced mandibular coefficient is $7{\cdot}66 \pm {\cdot}79$ which indicates distinct differentiation. The cranial types of these two series are very different, but it is suspected that this is due to the fact that the Hylam skulls were artificially deformed, though their facial and palatal measurements, which are not distinguished from the Fukien, do not appear to have been affected. The mandibles do distinguish the types, and this cannot be attributed to deformation.

These intra-group connexions do not encourage the hope that it will be possible to obtain any suggestive classification of the types from the reduced coefficients obtained from the mandibular measurements, and the comparison of series belonging to different groups makes this obvious. The most surprising connexion of the latter kind is the insignificant coefficient for the Farringdon Street and Punjabi series. But the Farringdon Street also has a lower value with the Tamil than with either the Anglo-Saxon or Dunstable series, and the Punjabi has a lower value with the Qau Egyptian than with any Asiatic series. It appears to be quite impossible to accept the coefficients as measures of racial relationship: they sometimes show close resemblance in type where no close racial affinity can be imagined, and they sometimes indicate clear distinction in type where close racial affinity must have existed. It may be noted that all the coefficients which make it impossible to obtain as suggestive a classification of the data as that obtained from the series previously dealt with are with one or other of three of the four new series. If the Spitalfields, Farringdon Street and Punjabi are omitted from Fig. 1 no connexions of the order considered are found between the three groups of series. The new material has apparently demonstrated the defect of the method.

It is clear that reduced coefficients of racial likeness for the mandible tend, in general, to be markedly lower than the values corresponding to them for the cranium, and only one case for which the reverse is true has been noted above, viz. that of the Anglo-Saxon and Farringdon Street series. This suggests that the unsatisfactory nature of the results shown in Fig. 1 may be due to the fact that some of the series used are too small for the purpose. The limiting size of cranial samples which yield suggestive and consistent results when compared in the same way has been determined empirically. In this case 50 is a safe limit to take, but samples composed of 30–50 individuals generally yield reliable results. The sizes of the series of mandibles can be judged from the $\overline{n}$'s in Table X, these being the average numbers of bones on which the means of the 10 characters used in computing the coefficients are based. Two of the male series have $\overline{n}$'s under 20, and no reliance whatever can be placed on results obtained from cranial samples no larger than these. If all series with $\overline{n}$'s less than 50 are ignored, we are only left with the Qau, Kerma and Gizeh (*E*) Egyptian and the Australian. The first

three are connected with one another and the last is widely removed from all of them by the coefficients, so there is nothing unexpected in these results, as far as they go. If the limit is lowered to 40, the Anglo-Saxon, Spitalfields and Punjabi series will also be included. Unexpected connexions are then found between the Punjabi series, on the one hand, and the Spitalfields, Gizeh Egyptian and Qau Egyptian, on the other, but of these three coefficients only one is less than 5 (Punjabi and Qau reduced $= 4{\cdot}50 \pm {\cdot}57$), and it has been shown that some ancient Egyptian and modern Indian cranial types are remarkably similar.* The results, which it is impossible to accept if the coefficients are considered as measures of racial relationship, are only evident when the shorter series are brought into the picture. The six insignificant coefficients, for example, are quite unexpected and unacceptable, but for every one of these one or both of the series compared has an $\overline{n}$ less than 40. The fact that several marked differences may be found between corresponding male and female coefficients in Table X, also suggests that several of the series are too short to give consistent results. The limiting size of sample required can only be determined empirically at present, as any theoretical estimate of it would require a knowledge of inter-racial variabilities which could only be found from far more extensive material than that available. It is quite possible that suggestive results would be given by series of mandibles made up by 50 or more individuals, or it may be necessary to adopt a still higher limiting size; and it may also be necessary to reject all reduced coefficients greater than 5, say, in interpreting such data. We cannot say that the method applied to measurements of series of mandibles is incapable of yielding results of value to the anthropologist, since it may be that the lack of suggestiveness of the arrangement shown in Fig. 1 is merely due to the fact that certain essential conditions were not observed in preparing that diagram. Data for additional series of a sufficient length will be required, either to justify the use of the coefficient of racial likeness in this case, or to demonstrate that it cannot be used profitably. All we can assert is that short series—composed of fewer than 40 mandibles, say—will not provide what is wanted.

Certain devices are used in calculating the reduced coefficients of racial likeness, and it may be suggested that these are partly responsible for the discordant results, and that the use of a theoretically more correct formula would modify them appreciably. The effect of the use of a single set of σ's instead of the values for each series used may be examined first. Martin has given reduced coefficients between the Egyptian E and a number of other series computed by using the Qau σ's, in one case, and those of the Egyptian E series itself, in the other.† Corresponding pairs of the lower coefficients are all in close agreement, though a few significant differences were found for the higher values which are

* See "A Study of the Badarian Crania recently excavated by the British School of Archaeology in Egypt", by Brenda N. Stoessiger, *Biometrika*, XIX (1927), pp. 110–50.

† *Loc. cit.* Table V.

neglected, however, in attempts to interpret the data. Three of the new male coefficients were calculated in the two ways with the following results, the first value of the reduced constant being that found by using the Qau σ's and the second that found by using the Egyptian E σ's: Farringdon Street and Punjabi $2{\cdot}15 \pm {\cdot}82$, $2{\cdot}56 \pm {\cdot}82$; Farringdon Street and Tamil $4{\cdot}94 \pm {\cdot}93$, $5{\cdot}21 \pm {\cdot}93$; Farringdon Street and Anglo-Saxon $17{\cdot}88 \pm {\cdot}83$, $22{\cdot}16 \pm {\cdot}83$. These results accord with Martin's, the coefficients calculated in the two ways being in close agreement in the case of the two lower pairs. It is unlikely that any of the unexpected relationships found between the series can be attributed to the use of a constant set of σ's in place of the sets for each of the series in a particular comparison.

It is unlikely, too, that the results obtained from the reduced coefficients of racial likeness differ appreciably from those which would be given by a theoretically more correct formula which takes into account the intra-racial correlations between the different measurements used. The 10 characters were chosen because the correlations between them are nearly all of a low order, and there is a far closer approach to the ideal condition here than in the case of the characters used in computing the cranial coefficients. Also, the mandibular coefficients which differ insignificantly from zero show a difference of means for nearly every character considered separately which would usually be considered insignificant,* and under these circumstances it cannot matter much whether the correlations between the characters are taken into account or not. These insignificant coefficients are largely responsible for our inability to accept the criterion as a measure of racial relationship.

Consideration of the same group makes it evident that the method of "reducing" the crude coefficient cannot be responsible for all the unexpected results obtained. In the case of the six crude coefficients which differ insignificantly from zero there is, in fact, no need to reduce them, and it is clear (*from cranial evidence*) that the device used achieves the end in view sufficiently well in other cases. As far as can be seen now, therefore, no one of the assumptions made, or devices used, in applying the method of the coefficient of racial likeness can be considered responsible for the failure of the method to give results of value. This failure may be due to the fact that it has been applied to samples which are too small. Another possibility is that the group of measurements used is unsuitable for the purpose in view, and this is discussed in the following section.

8. *A comparison of single characters.* The relative values of different characters for purposes of racial classification can be estimated from the α's found in computing the coefficients of racial likeness. An α is approximately the square of a quantity which is the difference of two means divided by its standard error, and the difference—if considered by itself—may be supposed clearly significant if the α is greater than 10. Comparisons have been made between 17 male series for the same 10 characters, so there is a total of 1360 α's for these. Of this total 374

* See p. 109 below.

(27·5 per cent.) are greater than 10, and for 12 of the 17 series Morant found a percentage of 27·3.* These percentages are in remarkably close agreement, in spite of the fact that they depend to a certain extent on the sizes of the series compared: on the average, comparisons of longer series will be expected to show more significant α's than comparisons of shorter series. But it is clear that for the series available the characters used are capable of making clear distinctions. For each of the 10 characters, in the comparison of the 17 series, there is a total of 136 α's. The percentages of α's greater than 10 are: $R\angle$ 12·5, rl 20·6, h_1 24·3, $c_y l$ 25·0, $m_2 p_1$ 25·7, zz 28·7, ml 28·7, rb' 30·9, $100\, g_o g_o / c_p l$ 36·8, w_1 41·9. Some characters evidently distinguish the types far more effectively than others, and it must be remembered that two were omitted from the list used in computing the coefficients because they appeared to be practically constant for the series considered by Morant. For the mandibular angle ($M\angle$) he only found three significant differences among 66 comparisons of mean values. This was the more surprising since anthropologists have often supposed that this character is of peculiar importance. It shows great intra-racial variability—the standard deviations for it being of the order 6°—and the means for 16 male series all lie between 120°·0 and 125°·3. It is true that the Australian mean of 117°·0 for 59 male mandibles is clearly divergent. The index expressing the breadth at the angles as a percentage of the breadth at the tips of the coronoid processes ($100\, g_o g_o / c_r c_r$) was also omitted because it only showed two significant differences in 66 comparisons. The intra-racial standard deviations for this character are of the order 7·0 and the range of the means for 17 male series is 95·0–102·9: the value of 100·4 for the Australian bones is not peculiar.

At the other extreme we find the mental angle ($C'\angle$) for which inter-racial variability is evidently much greater in proportion to intra-racial variability than in the case of the two preceding characters. This showed 35 significant differences among 66 comparisons of means, but it was not included among the characters used in computing coefficients of racial likeness because it was feared that it is a less reliable measurement than most of the others. It has been shown above (p. 89) that the readings of two observers sometimes show a very satisfactory agreement, and it is unlikely that personal equation is a disturbing factor in the case of comparisons between most of the means available for this character. A greater angle denotes a lesser projection of the chin. For 4 English male series the means range from 61°·8–70°·5, for 5 Egyptian from 70°·2–75°·5, for 7 Asiatic from 62°·9–77°·1 and the Australian mean for 40 bones is 78°·0. Several of these means are based on small numbers of specimens, and some have standard errors of the order 2°, but it is clear that the character often makes very clear distinctions between racial types. The Australian mean is extreme, but less removed from some of the others than would have been anticipated.

We may now ask whether the failure of the method of the coefficient of racial

* *Loc. cit.* p. 114.

likeness to give suggestive results when applied to series of mandibles is due to the choice of characters used in computing it. One of the most disconcerting results is the occurrence of insignificant coefficients in cases where distinct differentiation would have been expected, and where it is shown by the coefficients for the corresponding cranial series. Six insignificant values for male series of mandibles have been found (see Fig. 1), involving nine series. Standard deviations have only been given for four of these—the Dunstable and Farringdon Street English, the Punjabi and the Kerma Egyptian—as the others were considered too short for the purpose, and for the remaining five the Qau Egyptian constants may be applied, as in computing the coefficients. Comparisons of the means are summarized in Table XI, "none" signifying that there are no differences greater than three times their probable errors and the numbers in brackets being the ratios of the differences to their probable errors in cases where these are greater than 3.

TABLE XI

*A comparison of the significant differences between means for two groups of characters, in cases where the coefficients of racial likeness indicate an insignificant difference**

	10 C.B.L. characters	11 other characters
Anglo-Saxon and Dunstable	None	None
Farringdon Street and Punjabi	h_1 (4·4), $100\, g_o g_o / c_p l$ (3·6)	$g_o g_o$ (5·7), $C'\angle$ (5·2)
Sedment and Kerma Egyptian	ml (4·7)	$100\, c_r h/ml$ (3·4), $100\, c_r c_r/ml$ (3·2), $C'\angle$ (3·9)
Tamil and Nepalese	h_1 (3·1)	$C'\angle$ (4·6)
Tamil and Tibetan *A*	None	$C'\angle$ (5·5)
Nepalese and Tibetan *A*	None	$C'\angle$ (3·3)

* See text for explanation.

In these 6 cases the 11 characters which are not used in computing the coefficients do tend to show a larger number of significant differences, and clearer differentiation in the case of some characters, than do the 10 characters used. But this difference depends almost entirely on the mental angle ($C'\angle$), and if it were omitted the choice of any group of characters from the remaining 20 would lead to almost identically the same results as those derived from the group adopted: there would be no clear distinction between the pairs of series compared. The situation is changed if $C'\angle$ is included, but it is unsuitable as a coefficient of racial likeness character, since it is feared that its readings for some of the earlier series were not found in precisely the same way as that employed later.

The Australian series shows no low coefficient with any other, and this is largely due to the fact that two of its means (for zz and $m_2 p_1$) are the greatest yet found. But the same series also has the greatest $c_p l$ and $C'\angle$ and its $M\angle$ is the

smallest, and these three characters are not used in computing the coefficients. The type would almost certainly be distinguished equally clearly if the 11 remaining characters were used for this purpose instead of the 10 chosen. There is no doubt that the coefficients would be changed to some extent if they were computed for a different set of characters, but it seems probable that their orders would be little affected, and that the unexpected results which make it necessary to question the utility of the method would still be found.

A comparison of characters considered singly throws some light on the cause of these unexpected results. It will be sufficient to consider the 4 English series, which give mandibular coefficients markedly different from those found for the corresponding, but longer, series of crania. There are 6 comparisons, based on 10 characters, and there are only 20 of the 60 α's greater than 4. An α is approximately the square of a quantity which is the difference of two means divided by the standard error of the difference, and it would be expected to show some values greater than 4 in a set of 60 comparisons merely as the result of chance, if in fact all the series represented the same population. Only 7 of the α's are greater than 8, the largest being 22·9 and the next largest 17·9. For these four series there are very few differences which are markedly significant. There is sufficient evidence to show that some pairs of the types do differ significantly, but it is also clear that the estimates of divergence in type provided by the coefficients are likely to be particularly unreliable owing to the influence of errors of random sampling. For the material available errors of this kind may be large enough to obscure the situation. This seems to be a possibility, and in view of it our general conclusion must be not that coefficients of racial likeness based on measurements of series of mandibles are incapable of revealing racial relationships, but that longer series than some of those used above will be needed in order to examine the use of the method applied to such material. Short series—made up by fewer than 40 bones, say—will certainly not give what is needed.

9. *Conclusions.* This paper presents the results of a statistical treatment of two English (male and female), a Punjabi (male only) and an Australian (male and female) series of mandibles. Measurements were taken in accordance with the biometric technique, and estimates of their accuracy were obtained by repeating a number and comparing the distributions of first and second readings. The two English series are not associated with individual crania or other parts of the skeleton, and the problem of sexing these is discussed. It is shown that a crude mathematical method and anatomical appreciation agree in about 85 per cent. of cases, and there is reason to believe that the sexes finally adopted give the same order of accuracy as those obtained by sexing a series of crania anatomically. The constants of variation reveal a few significant, but very small absolute, differences between the variabilities of the different series available, and this conclusion is the same as that derived from cranial measurements. At the same time the mandible tends to be rather more variable, relative to size, than the

cranium. Special topics discussed are intra-racial correlations of the measurements, asymmetry and records relating to the teeth lost before death.

Racial comparisons are made by using the method of the coefficient of racial likeness. In all there are 17 male and 9 female series which can be used for this purpose, though several of these are evidently too small to be of any permanent value by themselves. The coefficients show a number of entirely unexpected resemblances and divergences, and it is clear that they do not provide a rational classification of the types. In general they differentiate the series far less effectively than do the corresponding cranial coefficients. In 6 cases out of 136 comparisons there is no evidence of a significant difference judging from the mandibular measurements, although the series would be expected to represent quite distinct races and the corresponding cranial coefficients indicate clear divergence. It is shown that this result is not due to an unsuitable choice of the characters used, and that it cannot be attributed, as far as can be seen, to the disturbing influence of any of the assumptions made, or devices used, in computing the coefficients. The failure of the method may well be due to the fact that short series of mandibles are not capable of providing a reliable classification of the races they represent. The longer series available do give suggestive results, but there are not enough of them to suggest that additional long series will probably do the same. We can assert that series made up by 40 or fewer individuals will not give the information required, and for such the lack of statistical distinction between two types cannot be supposed sufficient evidence of racial identity. Series made up by 40–50 individuals may be sufficiently long, but in further investigations on the same lines it would be safer to exclude all composed of fewer than 50 bones. It is quite likely that it will be possible to demonstrate the utility of the method when it is applied to sufficiently long series.

I wish to thank Dr Morant for the photographs reproduced, and Miss A. B. Clements for typing the manuscript of this paper.

DESCRIPTION OF PLATES

Plates I, II and III show standard aspects of typical male mandibles, the focal plane of the camera having been perpendicular or parallel to the standard horizontal plane of the bone. In these cases a lens with a long focal length was used, and the distance from lens to object was about 2½ metres. The small images obtained were enlarged in printing, and the prints are reproduced here approximately at 0·9 natural size. Distortion may be considered negligible. The photographs reproduced in Plates IV and V were taken with a lens having a shorter focal length and at a closer distance. The typical male mandibles were selected by considering the deviations of the measurements of shape (angles and indices) for each bone of a series from the means for the series in terms of the standard deviations for each of these measurements. Each of the three bones has every index and angle differing from the mean for the series to which it belongs by less than 1·2 times the standard deviation of the measurement. Also, their maximum breadths (bicondylar, w_1), lengths (total projective, ml) and heights (projective height of coronoid process, c_1h) fall within the same range, except that the coronoid height of the selected Farringdon Street mandible differs from the mean for the male series by an amount which is 1·7 times the standard deviation of the distribution. Bones which are more typical than the three shown could not be found in the short

series available, but it should be realized that a comparison of these may suggest that there are differences in metrical characters, or anatomical details, which would not be found if truly typical specimens—i.e. ones representing the averages in all respects—were available.

Plate I. Typical Punjabi (above, No. 6·3616) and Australian (No. 20·3003) male mandibles: *norma verticalis*. These two show little difference in size: for the true racial types the Australian seen from this aspect would show a rather larger excess in size over the Punjabi. There are clear differences in massiveness, and in the ways in which the teeth are set in the bones.

Plate II. Typical English (A, Farringdon Street, No. 622), Punjabi (B, No. 6·3616) and Australian (C, No. 20·3003) male mandibles: *norma lateralis*. The mandibular angles are seen to be very close and the means for the three series are closer still. For the true types the breadth of the ramus relative to its length would distinguish the Australian from the other two rather more clearly than is the case for the selected specimens. The lesser projection of the chin is the most striking characteristic of the Australian mandible, and this is also characteristic of the series.

Plate III. Typical English (A, Farringdon Street, No. 622), Punjabi (B, No. 6·3616) and Australian (C, No. 20·3003) male mandibles: *norma frontalis*. The differences in size are small, but the Australian is clearly the most massive bone and the setting of the teeth in it is characteristic.

Plate IV. Contrasted forms of Australian mandibles.

A. Male bones with extreme mental angles; 0·7 natural size. The mandible on the left (No. 20·59) has the lowest mental angle ($C' \angle = 65°{\cdot}5$) for the series, and the one on the right (No. 20·6213) the highest (92°·5). The mean angle is 78°·0 and the typical male (Plate II C) has a reading of 81°·5.

B. Female bones with extreme mental angles: 0·8 natural size. The mandible on the left (No. 20·6202) has the lowest mental angle (71°·0), and the one on the right (No. 20·8461) the highest (94°·0).

C. The dental arcades of two male Australian mandibles of contrasted forms: 0·9 natural size. The specimen on the left (No. 20·6211) has a parabolic arch and that on the right (No. 20·8521) differs from it in having the front teeth (incisors and canines) almost in a straight line. The difference is seen to be dependent more on the inclinations of the incisors than on the positions of their sockets.

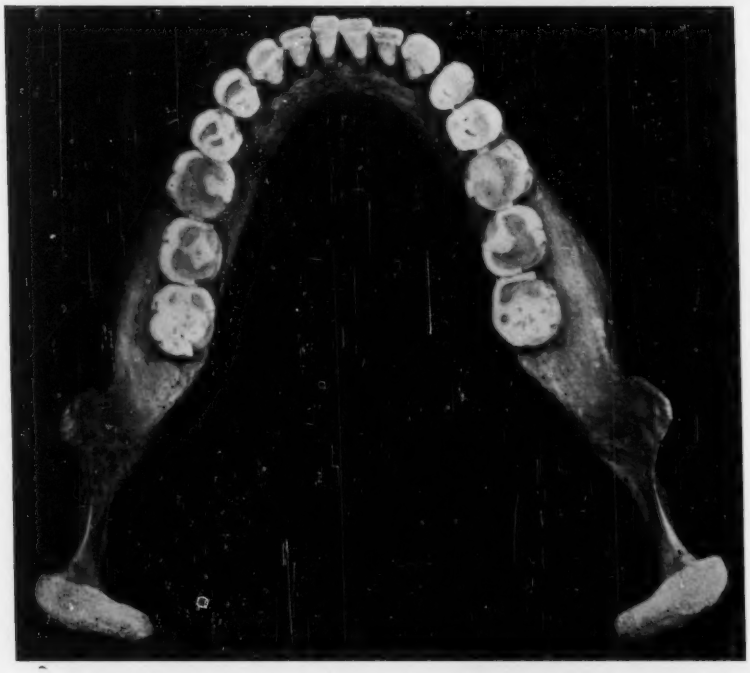
Plate V. Pathological and anomalous Australian mandibles.

A. A female mandible showing marked erosion of the angles due to syphilis: No. 3955·2, 0·9 natural size.

B. A male mandible showing severe healed injury of the right ramus: No. 20·8562, 0·9 natural size.

C. A male mandible of a remarkably massive and primitive type: No. 20·8551, 0·9 natural size.

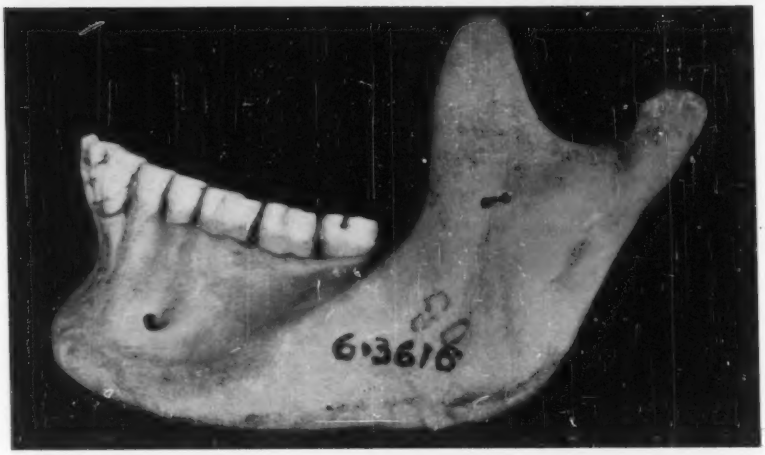
D. A male mandible showing gross overcrowding of the incisors: No. 20·7702, 1·3 natural size.



Typical Punjabi (above) and Australian Male Mandibles.

UNIV. OF MICH.

A

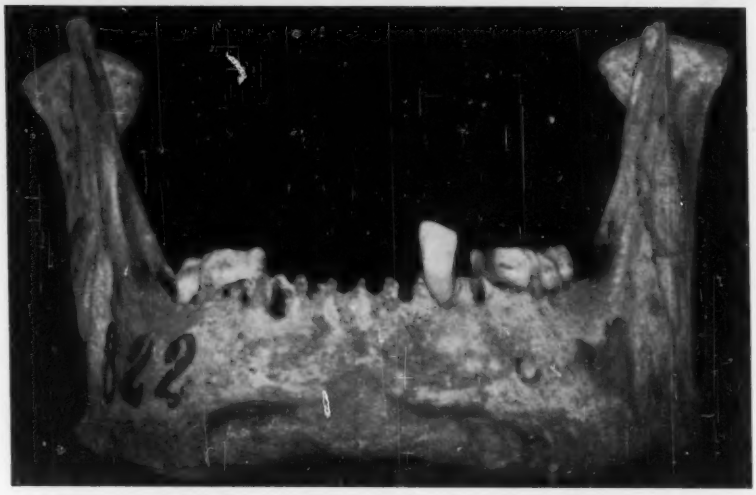


B

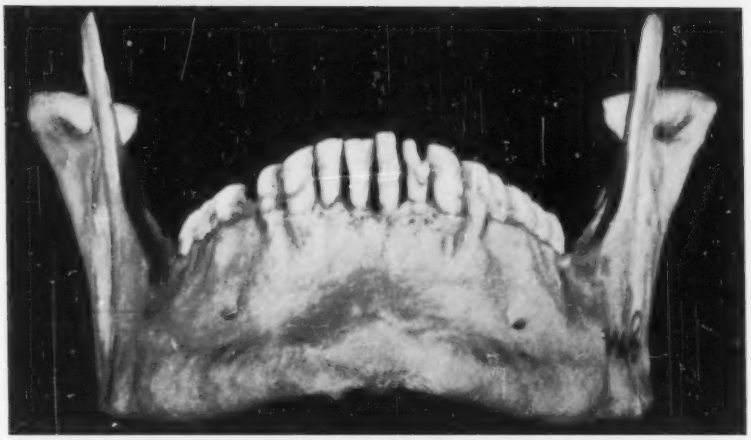
C

Typical English (A), Punjabi (B) and Australian (C) Male Mandibles.

UNIV. OF MICH.



A

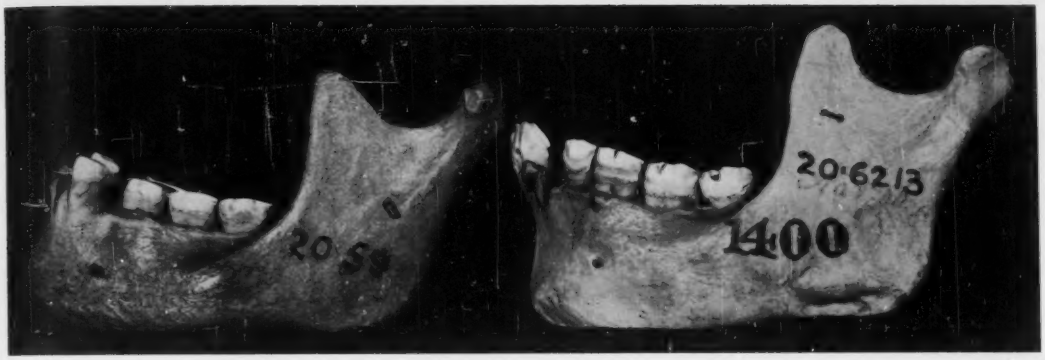


B

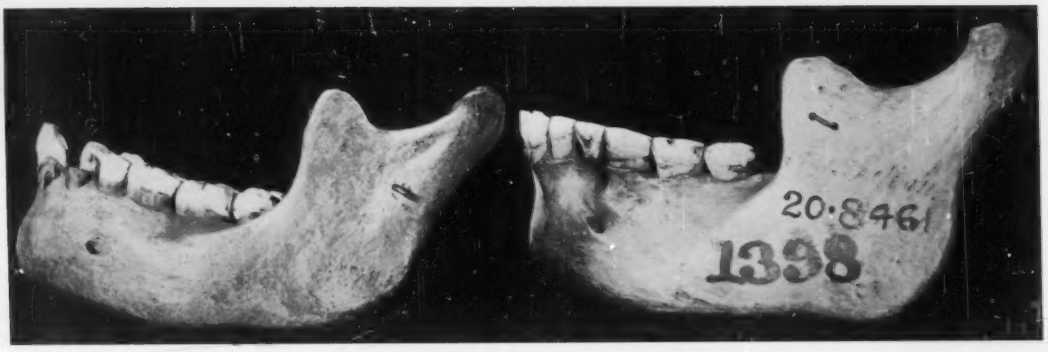
C

Typical English (A), Punjabi (B) and Australian (C) Male Mandibles.

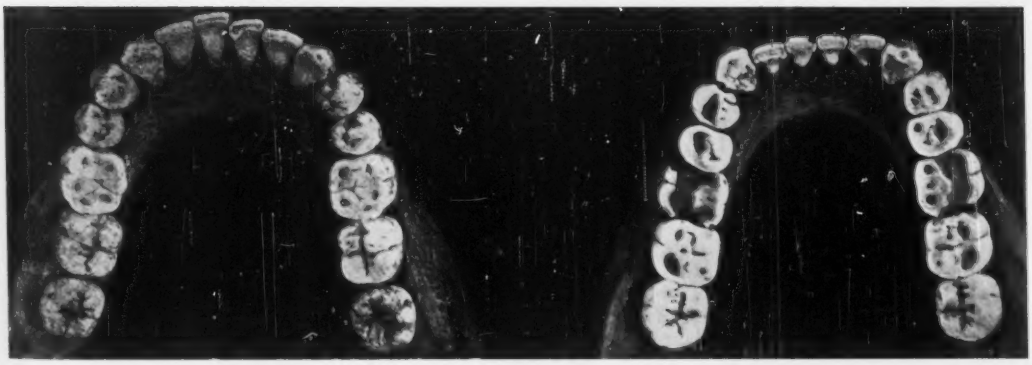
UNIV. OF



A. Male bones with extreme mental angles.



B. Female bones with extreme mental angles.



C. Different forms of the dental arcade.

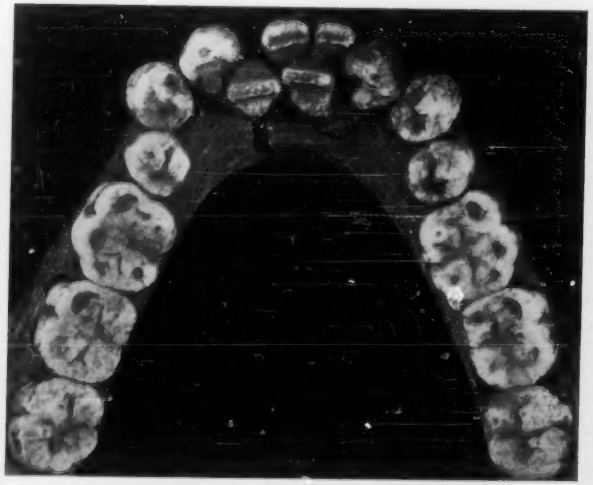
Contrasted Forms of Australian Mandibles.

UNIV. OF MICH.

A. Erosion of angles due to syphilis.

B. Healed fracture of right ramus.

C. A remarkably primitive type.



D. Gross overcrowding of incisors.

Pathological and Anomalous Australian Mandibles.

UNIV. OF MICH.

A BIOMETRIC STUDY OF THE HUMAN MALAR BONE

By T. L. WOO, Ph.D.

1. *Introduction.* The measurements of the skull which have been most widely used for anthropological purposes were originally defined, or elaborated from definitions of earlier workers, by Paul Broca and a number of his German contemporaries. The French and German techniques were by no means identical, but the two sets of measurements corresponded in a general way. They aimed at giving a general description of the cranium considered as a whole and of all its principal parts. The framers of the techniques were primarily anatomists who had become interested in anthropological problems, but there are no peculiarly anatomical considerations underlying their systems. In particular there was, in the case of the majority of the measurements, an unfortunate disregard of the fact that the skull is made up of a considerable number of different bones. Nearly all the later craniometric techniques are based on the earlier ones, and their general aim has been to secure greater precision and standardization. The result has been that a particular set of measurements has been given in a large number of publications for some tens of thousands of skulls representing extinct and existing races in all parts of the world. The value of this corpus of material is beyond question and, in fact, it is by far the most valuable material available at present which can be used to estimate with precision the resemblances of different varieties of man. It is known that all the usual measurements make some clear distinctions when the averages for different series are compared: in other words, they are all of racial significance. But it is also known that their relative values for the purpose of differentiating races differ greatly. Some appear to be almost constant for all races, while others usually show significant differences, and it is found that there is something like a gradual transition between these extremes. This grading of the characters, in order of their effectiveness as racial criteria, could not be appreciated until extensive data had been collected for them.

The position with regard to the customary measurements suggests that it should be possible to select a smaller number of characters which could be used as, or more, effectively for purposes of racial classification, with less labour involved in recording and computing. The list chosen might be made up partly by some of the old measurements and partly by new ones. It is generally recognized that certain features of the cranium which are obviously of value in aiding racial discrimination are not estimated by any of the classical measurements. New measurements of the "flatness" of the facial skeleton were taken on nearly 6000 skulls, representing a number of races from different parts of the

world, with the object of examining the value of one such feature.* It was concluded that a few of these are as useful for the purpose in view as any other characters that have been dealt with metrically, and far more useful than some for which extensive records are available.

The investigation described in the present paper was undertaken in the hope of discovering other new measurements which might be of exceptional value in aiding racial classification. In a paper published in 1931† the writer gave definitions of 25 chords and arcs, of which the majority were new, each being confined to a single bone of the skull. These were taken on a series—the Egyptian *E*—of nearly 900 male specimens obtained from a single cemetery at Gizeh which was used from the 26th to the 30th dynasties. Among the measurements dealt with in this study are two of the malar bone, the horizontal arc and the vertical arc. These were determined for both malar bones, so that the question of asymmetry could be investigated,‡ but no material was collected then to throw light on the possible sexual and racial significance of the measurements in question. One of the arcs was taken later by Dr von Bonin on a series of New Britain skulls,§ and the material for them presented below relates to an additional 710 crania, made up by 14 male and 2 female series representing races in different parts of the world. Two additional measurements of the malar bones of these 710 specimens were also recorded.

Anthropologists have hitherto devoted little attention to the malar bone, though a number of scattered remarks relating to racial differences in its size and form may be found in the literature. The earlier discussions of its metrical and anatomical variations are conveniently summarized by Le Double.‖ He remarks: "Il n'est pas démontré péremptoirement encore, par des mensurations multiples et précises, que le malaire ait, toutes choses égales d'ailleurs, des dimensions plus considérables dans une race que dans une autre et, dans une race quelconque, chez l'homme que chez la femme."

2. *The material measured.* All the skulls for which measurements of the malar bones are given for the first time in this paper are in the Museum of the Royal College of Surgeons, London. The writer measured them there in 1934 and he is greatly indebted to the authorities of the College, and particularly to Miss M. L. Tildesley, for granting him ready access to the specimens. The series are:

(i) *English.* 43 ♂. These came from a single cemetery at Portugal Street,

* T. L. Woo and G. M. Morant, "A Biometric Study of the 'Flatness' of the Facial Skeleton in Man", *Biometrika*, xxvi (1934), pp. 196–250.

† T. L. Woo, "On the Asymmetry of the Human Skull", *Ibid.* xxii, pp. 324–52.

‡ These malar bone measurements for the Egyptian series and the index derived from them are also treated by Karl Pearson and T. L. Woo in "Further Investigation of the Morphometric Characters of the Human Skull", *Ibid.* xxvii (1935), pp. 424–65.

§ "On the Craniology of Oceania. Crania from New Britain", *Ibid.* xxviii (1936), pp. 123–48.

‖ *Traité des Variations des Os de la Face de l'Homme* (1906), pp. 114–65.

London and they are known as the King's College series. The skulls are probably of eighteenth-century date.

(ii) *French.* 28♂. Most of these came from the catacombs of Paris and they are all later than the Merovingian period. Measurements were only taken of the complete crania.

(iii) *Italian.* 92♂. These modern skulls came from 12 provinces in the northern and central parts of Italy.

(iv) *Egyptian: dynastic.* 26♂. These belong to middle and late dynastic times. Measurements were only taken of the complete crania.

(v) *Egyptian: Ptolemaic and Roman.* 31♂. Measurements were only taken of the complete crania.

(vi) *Negro: Nigeria.* 41♂. Of this total 34 skulls came from South Nigeria —representing mainly the Ibibio and Ekoi tribes of the Calabar region—and the others are from different parts of North Nigeria.

(vii) *Negro: Congo.* 36♂ and 21♀. The majority of these specimens represent the Batetela tribe who live near the Lubefu River.

(viii) *Hindu: Bihar and Orissa.* 36♂. Several castes of Hindus are represented, and the majority of the specimens came from the Patna district in the north-west of Bihar.

(ix) *Punjabi.* 80♂. These crania are of Mohammedans and several castes of Hindus.

(x) *Javanese.* 45♂. These came from various parts of Java and the neighbouring islands.

(xi) *Chinese.* 63♂. Nearly half of these specimens are known to have come from various localities on the south-east coast of China, and the majority of the others probably came from the south of the country.

(xii) *Eskimo.* 29♂. These came from various parts of Greenland and neighbouring islands to the west.

(xiii) *Maori: New Zealand.* 39♂. The majority of these specimens came from the North Island, principally from the vicinity of Auckland, but some are from unknown localities.

(xiv) *Kanaka:* 50♂ and 50♀. These specimens came from the Islands of Oahu and Hawaii, and the population of the former is better represented than that of the latter.

Every one of these 639 male and 71 female skulls is sufficiently complete to give all the measurements defined in the following section.

3. *Definitions of measurements of the malar bone.* Fig. 1 shows the left malar bone and surrounding regions of the facial skeleton: *FMT* is the point where the malar ridge crosses the fronto-malar suture, and this is practically the same as Martin's *fronto-malare temporale*; *ZM* is the lowest point on the malar-maxillary suture, so it is his *zygomaxillare*. The other two points used are

not defined by Martin. *O* is the point where the malar-maxillary suture crosses the lower margin of the orbit, and *ZT* is the lowest point on the zygomatic suture which is still on the lateral surface of the arch. The measurements are:

(*a*) Ml_1 = minimum horizontal arc from *O* to *ZT*.

(*b*) Ml_2 = minimum vertical arc from *FMT* to *ZM*.

(*c*) $100\ Ml_2/Ml_1$.

These three are the malar-bone measurements of the earlier studies of measurements of single bones of the cranium. They are available for the long Egyptian series of male skulls* and for all the new material. Ml_1 is also available for the new British series. The arcs, taken with a steel tape, are recorded to the nearest 0·5 of a mm.

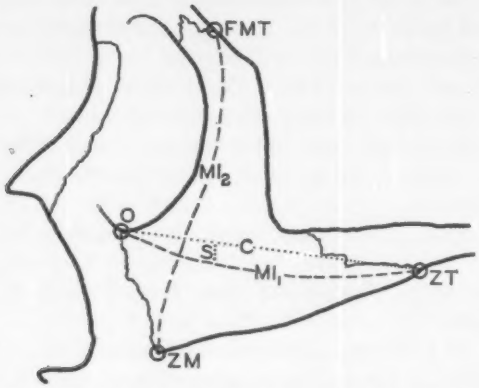


Fig. 1. The left malar bone and surrounding region, showing measurements taken.

(*d*) *C* = chord between the terminals of the horizontal arc (*O* and *ZT*).

(*e*) *S* = maximum subtense from this chord to the line marking the direction of the minimum horizontal arc (Ml_1). This line is first marked in pencil on the surface of the bone.

(*f*) $100\ S/C$. This provides a measure of the curvature of the horizontal arc.

These three measurements are only available for the new material.

The chord and the subtense were taken at the same time with the aid of a pair of co-ordinate callipers which was made for the writer by W. F. Stanley and Co. (London). This is similar in construction to the co-ordinate callipers made by P. Hermann, Rickenbach u. Sohn (Zürich), which could be used for the purpose, but the subtense arm of the new form terminates in a narrow straight

* An error was made in the tables of the asymmetry paper cited: the symbols Ml_1 and Ml_2 should be interchanged in these.

TABLE I. *Mean malar-bone measurements and their probable errors*

Series	Sex	No.	Horizontal arc (Ml_1)		Vertical arc (Ml_2)	
			R	*L*	*R*	*L*
Eskimo	♂	29	70·5 ± ·75	71·0 ± ·65	52·6 ± ·48	52·2 ± ·47
Chinese	♂	63	62·7 ± ·35	62·8 ± ·33	49·7 ± ·23	50·1 ± ·24
Javanese	♂	45	61·3 ± ·42	60·8 ± ·40	50·4 ± ·35	50·5 ± ·37
Kanaka	♂	50	62·3 ± ·34	62·5 ± ·36	49·7 ± ·25	49·7 ± ·25
Maori	♂	39	62·6 ± ·46	62·5 ± ·41	49·8 ± ·34	49·9 ± ·39
Negro: Nigeria	♂	41	60·8 ± ·46	61·6 ± ·49	49·2 ± ·29	49·4 ± ·33
Negro: Congo	♂	36	58·5 ± ·55	59·1 ± ·51	47·8 ± ·31	48·0 ± ·36
Egyptian: dynastic*	♂	26	58·3 ± ·60	58·6 ± ·49	47·3 ± ·41	47·7 ± ·41
Egyptian: 26th–30th dynasties	♂	716, etc.	59·4 ± ·110†	59·6 ± ·115†	49·4 ± ·073‡	50·0 ± ·076‡
Egyptian: Ptolemaic and Roman	♂	31	58·4 ± ·48	59·0 ± ·47	47·6 ± ·36	48·2 ± ·40
Hindu: Bihar and Orissa	♂	36	55·6 ± ·34	56·5 ± ·35	45·8 ± ·32	45·9 ± ·33
Punjabi	♂	80	57·6 ± ·33	58·2 ± ·33	47·3 ± ·20	47·5 ± ·25
Italian	♂	92	58·9 ± ·26	59·0 ± ·27	48·0 ± ·21	48·1 ± ·22
English	♂	43	58·8 ± ·40	59·1 ± ·39	49·8 ± ·25	50·2 ± ·30
French	♂	28	57·9 ± ·45	57·9 ± ·51	48·3 ± ·38	48·7 ± ·43
Kanaka	♀	50	57·9 ± ·36	58·5 ± ·38	46·2 ± ·25	45·9 ± ·26
Negro: Congo	♀	21	57·2 ± ·59	57·3 ± ·60	46·5 ± ·41	46·5 ± ·35

Series	Sex	No.	Horizontal chord (C)		Subtense to chord (S)	
			R	*L*	*R*	*L*
Eskimo	♂	29	62·4 ± ·59	62·1 ± ·59	14·5 ± ·24	14·7 ± ·30
Chinese	♂	63	55·3 ± ·30	54·9 ± ·26	12·5 ± ·14	12·5 ± ·14
Javanese	♂	45	54·3 ± ·40	53·9 ± ·36	12·1 ± ·17	11·9 ± ·17
Kanaka	♂	50	55·8 ± ·31	55·6 ± ·35	11·9 ± ·14	12·0 ± ·14
Maori	♂	39	56·0 ± ·39	55·3 ± ·37	11·7 ± ·11	11·8 ± ·16
Negro: Nigeria	♂	41	54·6 ± ·41	54·5 ± ·43	11·7 ± ·14	11·3 ± ·12
Negro: Congo	♂	36	52·1 ± ·44	52·3 ± ·46	10·8 ± ·16	10·6 ± ·15
Egyptian: dynastic*	♂	26	52·5 ± ·53	52·5 ± ·46	10·0 ± ·15	10·1 ± ·12
Egyptian: Ptolemaic and Roman	♂	31	52·5 ± ·36	53·0 ± ·39	10·9 ± ·17	10·9 ± ·18
Hindu: Bihar and Orissa	♂	36	49·8 ± ·30	50·1 ± ·29	10·1 ± ·12	10·2 ± ·10
Punjabi	♂	80	51·7 ± ·28	51·8 ± ·28	10·4 ± ·12	10·5 ± ·11
Italian	♂	92	53·5 ± ·24	53·2 ± ·25	9·6 ± ·08	9·8 ± ·09
English	♂	43	53·7 ± ·33	53·4 ± ·32	9·55 ± ·11	9·3 ± ·13
French	♂	28	52·9 ± ·38	52·5 ± ·39	9·3 ± ·15	9·2 ± ·15
Kanaka	♀	50	52·1 ± ·31	52·0 ± ·31	10·8 ± ·13	10·9 ± ·13
Negro: Congo	♀	21	51·2 ± ·51	50·8 ± ·50	10·6 ± ·21	10·5 ± ·22

* New Empire and late dynastic. † For 718 skulls. ‡ For 817 skulls. § For 716 s

I. *Mean malar-bone measurements and their probable err*

No.	Horizontal arc (Ml_1)		Vertical arc (Ml_2)	
	R	*L*	*R*	*L*
29	70·5 ± ·75	71·0 ± ·65	52·6 ± ·48	52·2 ±
63	62·7 ± ·35	62·8 ± ·33	49·7 ± ·23	50·1 ±
45	61·3 ± ·42	60·8 ± ·40	50·4 ± ·35	50·5 ±
50	62·3 ± ·34	62·5 ± ·36	49·7 ± ·25	49·7 ±
39	62·6 ± ·46	62·5 ± ·41	49·8 ± ·34	49·9 ±
41	60·8 ± ·46	61·6 ± ·49	49·2 ± ·29	49·4 ±
36	58·5 ± ·55	59·1 ± ·51	47·8 ± ·31	48·0 ±
26	58·3 ± ·60	58·6 ± ·49	47·3 ± ·41	47·7 ±
716, etc.	59·4 ± ·110†	59·6 ± ·115†	49·4 ± ·073‡	50·0 ±
31	58·4 ± ·48	59·0 ± ·47	47·6 ± ·36	48·2 ±
36	55·6 ± ·34	56·5 ± ·35	45·8 ± ·32	45·9 ±
80	57·6 ± ·33	58·2 ± ·33	47·3 ± ·20	47·5 ±
92	58·9 ± ·26	59·0 ± ·27	48·0 ± ·21	48·1 ±
43	58·8 ± ·40	59·1 ± ·39	49·8 ± ·25	50·2 ±
28	57·9 ± ·45	57·9 ± ·51	48·3 ± ·38	48·7 ±
50	57·9 ± ·36	58·5 ± ·38	46·2 ± ·25	45·9 ±
21	57·2 ± ·59	57·3 ± ·60	46·5 ± ·41	46·5 ±

No.	Horizontal chord (C)		Subtense to chord (S	
	R	*L*	*R*	*L*
29	62·4 ± ·59	62·1 ± ·59	14·5 ± ·24	14·7 ±
63	55·3 ± ·30	54·9 ± ·26	12·5 ± ·14	12·5 ±
45	54·3 ± ·40	53·9 ± ·36	12·1 ± ·17	11·9 ±
50	55·8 ± ·31	55·6 ± ·35	11·9 ± ·14	12·0 ±
39	56·0 ± ·39	55·3 ± ·37	11·7 ± ·11	11·8 ±
41	54·6 ± ·41	54·5 ± ·43	11·7 ± ·14	11·3 ±
36	52·1 ± ·44	52·3 ± ·46	10·8 ± ·16	10·6 ±
26	52·5 ± ·53	52·5 ± ·46	10·0 ± ·15	10·1 ±
31	52·5 ± ·36	53·0 ± ·39	10·9 ± ·17	10·9 ±
36	49·8 ± ·30	50·1 ± ·29	10·1 ± ·12	10·2 ±
80	51·7 ± ·28	51·8 ± ·28	10·4 ± ·12	10·5 ±
92	53·5 ± ·24	53·2 ± ·25	9·6 ± ·08	9·8 ±
43	53·7 ± ·33	53·4 ± ·32	9·55 ± ·11	9·3 ±
28	52·9 ± ·38	52·5 ± ·39	9·3 ± ·15	9·2 ±
50	52·1 ± ·31	52·0 ± ·31	10·8 ± ·13	10·9 ±
21	51·2 ± ·51	50·8 ± ·50	10·6 ± ·21	10·5 ±

nd late dynastic. † For 718 skulls. ‡ For 817 skulls. § For

ir probable errors

rtical arc (Ml_2)		100 Ml_2/Ml_1	
	L	*R*	*L*
·48	52·2 ± ·47	74·8 ± ·69	73·4 ± ·73
·23	50·1 ± ·24	79·4 ± ·53	80·0 ± ·51
·35	50·5 ± ·37	82·4 ± ·58	83·3 ± ·69
·25	49·7 ± ·25	79·8 ± ·47	79·6 ± ·49
·34	49·9 ± ·39	79·7 ± ·57	80·0 ± ·60
·29	49·4 ± ·33	81·3 ± ·59	80·4 ± ·66
·31	48·0 ± ·36	82·2 ± ·69	81·1 ± ·71
·41	47·7 ± ·41	81·3 ± ·69	81·9 ± ·70
·073‡	50·0 ± ·076‡	83·2 ± ·148§	84·0 ± ·152§
·36	48·2 ± ·40	81·9 ± ·71	82·5 ± ·78
·32	45·9 ± ·33	82·4 ± ·62	81·4 ± ·59
·20	47·5 ± ·25	82·5 ± ·49	81·9 ± ·49
·21	48·1 ± ·22	81·7 ± ·38	81·8 ± ·40
·25	50·2 ± ·30	85·0 ± ·43	85·3 ± ·52
·38	48·7 ± ·43	83·6 ± ·81	84·5 ± ·80
·25	45·9 ± ·26	79·8 ± ·37	78·7 ± ·44
·41	46·5 ± ·35	81·6 ± ·83	81·5 ± ·80

tense to chord (S)		100 S/C	
	L	*R*	*L*
·24	14·7 ± ·30	23·2 ± ·26	23·6 ± ·29
·14	12·5 ± ·14	22·5 ± ·21	22·7 ± ·23
·17	11·9 ± ·17	22·2 ± ·26	22·0 ± ·25
·14	12·0 ± ·14	21·4 ± ·21	21·7 ± ·23
·11	11·8 ± ·16	20·9 ± ·20	21·3 ± ·24
·14	11·3 ± ·12	21·4 ± ·23	20·8 ± ·20
·16	10·6 ± ·15	20·7 ± ·23	20·3 ± ·21
·15	10·1 ± ·12	19·1 ± ·25	19·3 ± ·24
·17	10·9 ± ·18	20·9 ± ·26	20·5 ± ·27
·12	10·2 ± ·10	20·4 ± ·24	20·4 ± ·20
·12	10·5 ± ·11	20·1 ± ·18	20·1 ± ·17
·08	9·8 ± ·09	18·0 ± ·14	18·4 ± ·14
± ·11	9·3 ± ·13	17·8 ± ·23	17·5 ± ·21
·15	9·2 ± ·15	17·7 ± ·25	17·4 ± ·20
·13	10·9 ± ·13	20·7 ± ·22	20·8 ± ·19
·21	10·5 ± ·22	20·7 ± ·35	20·7 ± ·35

skulls. § For 716 skulls.

TABLE II. *Variabili*

Series	Sex	No.	Horizontal arc (*M*		
			σ		Coef
			R	*L*	*R*
Eskimo	♂	29	6·00 ± ·53	5·21 ± ·46	8·51 ±
Chinese	♂	63	4·07 ± ·25	3·86 ± ·23	6·49 ±
Javanese	♂	45	4·18 ± ·30	4·00 ± ·28	6·82 ±
Kanaka	♂	50	3·52 ± ·24	3·74 ± ·25	5·65 ±
Maori	♂	39	4·29 ± ·33	3·83 ± ·29	6·86 ±
Negro: Nigeria	♂	41	4·40 ± ·33	4·65 ± ·35	7·24 ±
Negro: Congo	♂	36	4·89 ± ·39	4·55 ± ·36	8·35 ±
Egyptian: dynastic*	♂	26	4·53 ± ·42	3·71 ± ·35	7·76 ±
Egyptian: 20th–30th dynasties	♂	716, etc.†	4·38 ± ·08	4·55 ± ·08	7·37 ±
Egyptian: Ptolemaic and Roman	♂	31	3·99 ± ·34	3·90 ± ·33	6·83 ±
Hindu: Bihar and Orissa	♂	36	2·99 ± ·24	3·09 ± ·25	5·38 ±
Punjabi	♂	80	4·32 ± ·23	4·37 ± ·23	7·51 ±
Italian	♂	92	3·70 ± ·18	3·83 ± ·19	6·28 ±
English	♂	43	3·91 ± ·29	3·76 ± ·27	6·65 ±
French	♂	28	3·55 ± ·32	3·99 ± ·36	6·14 ±
Kanaka	♀	50	3·75 ± ·25	3·94 ± ·27	6·47 ±
Negro: Congo	♀	21	4·02 ± ·42	4·10 ± ·43	7·03 ±

Series	Sex	No.	Horizontal chord		
			σ		Coef
			R	*L*	*R*
Eskimo	♂	29	4·72 ± ·42	4·74 ± ·42	7·56 ±
Chinese	♂	63	3·50 ± ·21	3·10 ± ·19	6·32 ±
Javanese	♂	45	3·99 ± ·28	3·56 ± ·25	7·34 ±
Kanaka	♂	50	3·21 ± ·22	3·69 ± ·25	5·76 ±
Maori	♂	39	3·62 ± ·28	3·39 ± ·26	6·47 ±
Negro: Nigeria	♂	41	3·86 ± ·29	4·04 ± ·31	7·07 ±
Negro: Congo	♂	36	3·96 ± ·31	4·09 ± ·33	7·60 ±
Egyptian: dynastic*	♂	26	3·98 ± ·37	3·46 ± ·32	7·58 ±
Egyptian: Ptolemaic and Roman	♂	31	3·00 ± ·26	3·20 ± ·27	5·71 ±
Hindu: Bihar and Orissa	♂	36	2·70 ± ·22	2·56 ± ·20	5·43 ±
Punjabi	♂	80	3·72 ± ·20	3·76 ± ·20	7·19 ±
Italian	♂	92	3·37 ± ·17	3·50 ± ·17	6·31 ±
English	♂	43	3·19 ± ·23	3·08 ± ·22	5·93 ±
French	♂	28	3·00 ± ·27	3·05 ± ·27	5·67 ±
Kanaka	♀	50	3·27 ± ·22	3·24 ± ·22	6·27 ±
Negro: Congo	♀	21	3·45 ± ·36	3·42 ± ·36	6·75 ±

* New Empire and late dy

I. *Variabilities of malar-bone measurements*

Horizontal arc (Ml_1)			Vertical a		
	Coefficient of variation		σ		
L	*R*	*L*	*R*	*L*	
5·21 ± ·46	8·51 ± ·76	7·33 ± ·65	3·81 ± ·34	3·73 ± ·33	7
3·86 ± ·23	6·49 ± ·39	6·14 ± ·37	2·66 ± ·16	2·82 ± ·17	5
4·00 ± ·28	6·82 ± ·49	6·58 ± ·47	3·48 ± ·25	3·66 ± ·26	6
3·74 ± ·25	5·65 ± ·38	5·98 ± ·40	2·65 ± ·18	2·66 ± ·18	5
3·83 ± ·29	6·86 ± ·53	6·13 ± ·47	3·17 ± ·24	3·61 ± ·28	6
4·65 ± ·35	7·24 ± ·54	7·55 ± ·57	2·76 ± ·21	3·10 ± ·23	5
4·55 ± ·36	8·35 ± ·67	7·70 ± ·62	2·77 ± ·22	3·18 ± ·25	5
3·71 ± ·35	7·76 ± ·73	6·32 ± ·59	3·10 ± ·29	3·11 ± ·29	6
4·55 ± ·08	7·37 ± ·13	7·64 ± ·14	3·11 ± ·05	3·21 ± ·05	6
3·90 ± ·33	6·83 ± ·59	6·61 ± ·57	3·01 ± ·26	3·28 ± ·28	6
3·09 ± ·25	5·38 ± ·43	5·47 ± ·44	2·87 ± ·23	2·97 ± ·24	6
4·37 ± ·23	7·51 ± ·40	7·51 ± ·40	2·70 ± ·14	3·34 ± ·18	5
3·83 ± ·19	6·28 ± ·31	6·49 ± ·32	2·95 ± ·15	3·07 ± ·15	6
3·76 ± ·27	6·65 ± ·49	6·36 ± ·46	2·40 ± ·18	2·96 ± ·22	4
3·99 ± ·36	6·14 ± ·56	6·90 ± ·62	3·01 ± ·27	3·40 ± ·31	6
3·94 ± ·27	6·47 ± ·44	6·73 ± ·46	2·62 ± ·18	2·69 ± ·18	5
4·10 ± ·43	7·03 ± ·74	7·16 ± ·75	2·77 ± ·29	2·40 ± ·25	5

Horizontal chord (C)			Subtense to		
	Coefficient of variation		σ		
L	*R*	*L*	*R*	*L*	
4·74 ± ·42	7·56 ± ·67	7·63 ± ·68	1·89 ± ·17	2·36 ± ·21	13
3·10 ± ·19	6·32 ± ·38	5·64 ± ·34	1·61 ± ·10	1·70 ± ·10	12
3·56 ± ·25	7·34 ± ·52	6·60 ± ·47	1·67 ± ·12	1·66 ± ·12	13
3·69 ± ·25	5·76 ± ·39	6·64 ± ·45	1·43 ± ·10	1·51 ± ·10	12
3·39 ± ·26	6·47 ± ·50	6·13 ± ·47	1·04 ± ·08	1·47 ± ·11	8
4·04 ± ·31	7·07 ± ·53	7·40 ± ·55	1·36 ± ·10	1·16 ± ·09	11
4·09 ± ·33	7·60 ± ·61	7·82 ± ·63	1·44 ± ·11	1·36 ± ·11	13
3·46 ± ·32	7·58 ± ·71	6·60 ± ·62	1·17 ± ·11	0·89 ± ·08	11
3·20 ± ·27	5·71 ± ·49	6·05 ± ·52	1·36 ± ·12	1·49 ± ·13	12
2·56 ± ·20	5·43 ± ·43	5·11 ± ·41	1·10 ± ·09	0·93 ± ·07	10
3·76 ± ·20	7·19 ± ·39	7·26 ± ·39	1·57 ± ·08	1·48 ± ·08	15
3·50 ± ·17	6·31 ± ·31	6·58 ± ·33	1·20 ± ·06	1·32 ± ·07	12
3·08 ± ·22	5·93 ± ·43	5·78 ± ·42	1·27 ± ·09	1·05 ± ·08	13
3·05 ± ·27	5·67 ± ·51	5·81 ± ·53	1·17 ± ·11	1·15 ± ·10	12
3·24 ± ·22	6·27 ± ·42	6·22 ± ·42	1·39 ± ·09	1·37 ± ·09	12
3·42 ± ·36	6·75 ± ·71	6·74 ± ·70	1·45 ± ·15	1·52 ± ·16	13

mpire and late dynastic.

† See Table I.

ABLE II. *Variabilities of malar-bone measurements*

Horizontal arc (Ml_1)				Vertical arc (Ml_2)				$100\ Ml_2/Ml_1$	
σ		Coefficient of variation		σ		Coefficient of variation		σ	
R	*L*	*R*	*L*	*R*	*L*	*R*	*L*	*R*	*L*
00 ±·53	5·21 ±·46	8·51 ±·76	7·33 ±·65	3·81 ±·34	3·73 ±·33	7·25 ±0·65	7·15 ±0·64	5·54 ±·49	5·80 ±·51
07 ±·25	3·86 ±·23	6·49 ±·39	6·14 ±·37	2·66 ±·16	2·82 ±·17	5·36 ±0·32	5·62 ±0·34	6·29 ±·38	5·95 ±·36
18 ±·30	4·00 ±·28	6·82 ±·49	6·58 ±·47	3·48 ±·25	3·66 ±·26	6·91 ±0·49	7·26 ±0·52	5·80 ±·41	6·83 ±·49
52 ±·24	3·74 ±·25	5·65 ±·38	5·98 ±·40	2·65 ±·18	2·66 ±·18	5·32 ±0·36	5·36 ±0·36	4·96 ±·33	5·18 ±·35
29 ±·33	3·83 ±·29	6·86 ±·53	6·13 ±·47	3·17 ±·24	3·61 ±·28	6·36 ±0·49	7·24 ±0·56	5·31 ±·41	5·60 ±·43
40 ±·33	4·65 ±·35	7·24 ±·54	7·55 ±·57	2·76 ±·21	3·10 ±·23	5·61 ±0·42	6·28 ±0·47	5·58 ±·42	6·22 ±·46
89 ±·39	4·55 ±·36	8·35 ±·67	7·70 ±·62	2·77 ±·22	3·18 ±·25	5·80 ±0·46	6·62 ±0·53	6·13 ±·49	6·35 ±·50
53 ±·42	3·71 ±·35	7·76 ±·73	6·32 ±·59	3·10 ±·29	3·11 ±·29	6·56 ±0·62	6·52 ±0·61	5·18 ±·48	5·25 ±·49
38 ±·08	4·55 ±·08	7·37 ±·13	7·64 ±·14	3·11 ±·05	3·21 ±·05	6·30 ±0·11	6·42 ±0·11	5·85 ±·10	6·02 ±·11
99 ±·34	3·90 ±·33	6·83 ±·59	6·61 ±·57	3·01 ±·26	3·28 ±·28	6·32 ±0·54	6·79 ±0·58	5·86 ±·50	6·40 ±·55
99 ±·24	3·09 ±·25	5·38 ±·43	5·47 ±·44	2·87 ±·23	2·97 ±·24	6·26 ±0·50	6·47 ±0·52	5·53 ±·44	5·23 ±·42
32 ±·23	4·37 ±·23	7·51 ±·40	7·51 ±·40	2·70 ±·14	3·34 ±·18	5·70 ±0·30	7·04 ±0·38	6·56 ±·35	6·55 ±·35
70 ±·18	3·83 ±·19	6·28 ±·31	6·49 ±·32	2·95 ±·15	3·07 ±·15	6·15 ±0·31	6·38 ±0·32	5·40 ±·27	5·64 ±·28
91 ±·29	3·76 ±·27	6·65 ±·49	6·36 ±·46	2·40 ±·18	2·96 ±·22	4·82 ±0·35	5·39 ±0·43	4·18 ±·30	5·02 ±·36
55 ±·32	3·99 ±·36	6·14 ±·56	6·90 ±·62	3·01 ±·27	3·40 ±·31	6·23 ±0·56	6·98 ±0·63	6·39 ±·58	6·28 ±·57
75 ±·25	3·94 ±·27	6·47 ±·44	6·73 ±·46	2·62 ±·18	2·69 ±·18	5·68 ±0·38	5·86 ±0·40	3·92 ±·26	4·56 ±·31
02 ±·42	4·10 ±·43	7·03 ±·74	7·16 ±·75	2·77 ±·29	2·40 ±·25	5·96 ±0·62	5·16 ±0·54	5·62 ±·59	5·40 ±·56

Horizontal chord (C)				Subtense to chord (S)				$100\ S/C$	
σ		Coefficient of variation		σ		Coefficient of variation		σ	
R	*L*	*R*	*L*	*R*	*L*	*R*	*L*	*R*	*L*
72 ±·42	4·74 ±·42	7·56 ±·67	7·63 ±·68	1·89 ±·17	2·36 ±·21	13·05 ±1·18	16·03 ±1·46	2·09 ±·19	2·30 ±·20
50 ±·21	3·10 ±·19	6·32 ±·38	5·64 ±·34	1·61 ±·10	1·70 ±·10	12·84 ±0·78	13·60 ±0·83	2·41 ±·15	2·66 ±·16
99 ±·28	3·56 ±·25	7·34 ±·52	6·60 ±·47	1·67 ±·12	1·66 ±·12	13·72 ±0·99	14·06 ±1·02	2·54 ±·18	2·47 ±·18
21 ±·22	3·69 ±·25	5·76 ±·39	6·64 ±·45	1·43 ±·10	1·51 ±·10	12·03 ±0·82	12·55 ±0·86	2·15 ±·15	2·41 ±·16
32 ±·28	3·39 ±·26	6·47 ±·50	6·13 ±·47	1·04 ±·08	1·47 ±·11	8·86 ±0·68	12·48 ±0·97	1·86 ±·14	2·22 ±·17
86 ±·29	4·04 ±·31	7·07 ±·53	7·40 ±·55	1·36 ±·10	1·16 ±·09	11·64 ±0·88	10·27 ±0·77	2·14 ±·16	1·88 ±·14
96 ±·31	4·09 ±·33	7·60 ±·61	7 82 ±·63	1·44 ±·11	1·36 ±·11	13·31 ±1·08	12·84 ±1·04	2·02 ±·16	1·91 ±·15
98 ±·37	3·46 ±·32	7·58 ±·71	6·60 ±·62	1·17 ±·11	0·89 ±·08	11·67 ±1·11	8·78 ±0·83	1·86 ±·17	1·80 ±·17
00 ±·26	3·20 ±·27	5·71 ±·49	6·05 ±·52	1·36 ±·12	1·49 ±·13	12·50 ±1·09	13·67 ±1·19	2·14 ±·18	2·20 ±·19
70 ±·22	2·56 ±·20	5·43 ±·43	5·11 ±·41	1·10 ±·09	0·93 ±·07	10·84 ±0·87	9·08 ±0·73	2·10 ±·17	1·75 ±·14
72 ±·20	3·76 ±·20	7·19 ±·39	7·26 ±·39	1·57 ±·08	1·48 ±·08	15·12 ±0·82	14·14 ±0·77	2·39 ±·13	2·25 ±·12
37 ±·17	3·50 ±·17	6·31 ±·31	6·58 ±·33	1·20 ±·06	1·32 ±·07	12·49 ±0·63	12·28 ±0·62	2·02 ±·10	2·04 ±·10
19 ±·23	3·08 ±·22	5·93 ±·43	5·78 ±·42	1·27 ±·09	1·05 ±·08	13·33 ±0·99	11·26 ±0·83	2·21 ±·16	2·00 ±·15
00 ±·27	3·05 ±·27	5·67 ±·51	5·81 ±·53	1·17 ±·11	1·15 ±·10	12·57 ±1·15	12·48 ±1·14	1·97 ±·18	1·57 ±·14
27 ±·22	3·24 ±·22	6·27 ±·42	6·22 ±·42	1·39 ±·09	1·37 ±·09	12·83 ±0·88	12·57 ±0·86	2·30 ±·16	2·03 ±·14
45 ±·36	3·42 ±·36	6·75 ±·71	6·74 ±·70	1·45 ±·15	1·52 ±·16	13·66 ±1·45	14·52 ±1·54	2·35 ±·25	2·39 ±·25

* New Empire and late dynastic.

† See Table I.

edge instead of a sharp tip. This makes it possible to determine the maximum subtense by merely bringing the edge down on to the arc, marked in pencil on the bone, from which the subtense is taken: there is no need to repeat this operation several times in order to find the maximum reading, as is necessary when the form of callipers with a pointed subtense arm is used. Both scales of the new instrument have verniers attached, and the chords and subtenses were recorded to the nearest 0·1 of a mm.

4. *Sexual and bilateral comparisons.* The metrical material available for estimating sexual differences for the malar bone is very scanty. There is one series of 50 male and 50 female Kanaka skulls and another of 36 male and 21 female Congo-Negro. Means are given in Table I and variabilities in Table II. The numbers are too small to give reliable sex ratios (male mean/female mean) for the absolute measurements, but as far as can be seen these are not peculiar for cranial measurements. The eight constants range from 1·068 to 1·102 for the Kanaka series and from 1·010 to 1·032 for the Congo, but it would be unwise to assume from such slender evidence that the races are distinguished by their average sex differences. There is no suggestion that the true ratios are significantly different for different measurements, or for the right and left sides in the case of the same measurement. No clearly significant differences are found between the corresponding male and female indices, and it is clear that any sexual differentiation—apart from that in absolute size—which might be deduced from the measurements could only be revealed by data for larger samples.

The data which can be used to examine bilateral differences for male skulls are far more extensive. In estimating the significance of such constants the bilateral correlations have to be taken into account, and these are given in Table III for the long Egyptian and the two longest of the new series. In the case of the three comparisons which can be made, the Egyptian constant is greater than those for the other two series, and most of the amounts by which its value exceeds theirs are significant. For all six characters, however, no significant differences are found between the corresponding Italian and Punjabi correlations. It is commonly found for anthropometric material that the more homogeneous series give the higher correlations, and this relation is observed in the present case. There are no marked differences between the correlations for different characters, except that those for the subtense and the index involving the subtense tend to be lower than the others. It may be suggested that this is due to the fact that readings of the subtense—quite the smallest measurement—were not recorded in sufficiently small units to give reliable correlations. An examination of the constants for the Italian series shows that this is not the case, however. Its highest bilateral correlation is for the vertical arc. Readings of this measurement were taken to the nearest 0·5 of a mm. and the standard deviation for it is about 3 mm., so the unit of measurement is about one-sixth

of the σ. Readings of the subtense were taken to the nearest 0·1 of a mm. and the standard deviation for it is of the order 1·2 mm., so the unit of measurement is here about one-twelfth of the σ. Merely on account of the way in which the measurements were taken, it might hence be expected that the arc would show a lower bilateral correlation than the subtense, but actually the latter has the lower value.

Comparisons between the means and standard deviations for the right and left sides are summarized in Table IV, and the data there refer only to the fourteen new series. Treating each character separately, the numbers of series showing the right constant greater than the left, equality, and the left constant greater than the right are given, and also all the ratios of the differences to their probable errors which exceed 3·5. In calculating these ratios* the bilateral correlations used were: (i) in the case of Ml_1, Ml_2 and $100\ Ml_2/Ml_1$ those of the long Egyptian series in all comparisons except those for the Italian and Punjabi series, the appropriate correlations in Table III being used for these; (ii) in the case of C, S and $100\ S/C$, those of the Italian series in all comparisons except the Punjabi. It will be seen from Table IV that few markedly significant differences are found for any character: larger series than any dealt with there are generally needed to reveal beyond question the asymmetry in type of any cranial measurement. Considering all the series together, there is a clear suggestion that both the horizontal and vertical arcs of the malar bone tend to be larger in size on the left than on the right. This accords with the results obtained for the paired bones—approximately 800 in number—of the male Egyptian series E skulls for which both differences of means are significant and of the same sign ($L > R$). For the 50 male skulls from New Britain Dr von Bonin found the left mean of the horizontal arc 0·1 mm. greater than the right, though this difference is quite insignificant. The means of the other characters give no clear indication of asymmetry and it is curious that this should be so for the horizontal chord, since the arc which has the same terminals shows a different relation. The comparisons in Table IV for standard deviations suggest that variability on the left side exceeds that on the right in the case of Ml_2 and $100\,Ml_2/Ml_1$, but that there is no bilateral difference in variability in the case of the other characters. For the long Egyptian series the left standard deviation was found to be significantly in excess of the right in the case of both Ml_1 and Ml_2. Of the four absolute measurements, Ml_2 is the only one for which there is a suggestion of a bilateral difference in relative variability, measured by the coefficient of variation. For this character the left constant exceeds the right in the case of 12 of the 14 short series: the long Egyptian series shows differences of the same sign for both Ml_1 and Ml_2, but the former is significant and the latter is not.

* The formulae which have to be used are given on pp. 329 and 337 of the writer's paper on asymmetry cited above.

TABLE III. *Bilateral correlations of measurements of malar bones: male series*

	No.	Horizontal arc (Ml_1)	Vertical arc (Ml_2)	Horizontal chord (C)
Italian	92	·7327 ± ·0223	·9123 ± ·0118	·8708 ± ·0170
Punjabi	80	·8248 ± ·0241	·8498 ± ·0210	·9130 ± ·0126
Egyptian: 26th–30th dynasties	716, etc.	·9399 ± ·0029*	·9219 ± ·0035†	—

	No.	Subtense to chord (S)	100 Ml_2/Ml_1	100 S/C
Italian	92	·6644 ± ·0393	·8017 ± ·0251	·5782 ± ·0468
Punjabi	80	·7541 ± ·0325	·7430 ± ·0338	·6970 ± ·0366
Egyptian: 26th–30th dynasties	716, etc.	—	·8806 ± ·0057‡	—

* For 718 skulls. † For 817 skulls. ‡ For 716 skulls.

TABLE IV. *Bilateral comparisons of constants for malar-bone measurements: fourteen male series*

	$R>L$		Equality	$L>R$	
	No. of cases	Significant differences	No. of cases	No. of cases	Significant differences
	Means				
Horizontal arc (Ml_1)	2	—	1	11	Negro: Nigeria (4·8), Egyptian: Ptolemaic and Roman (3·6), Hindu (7·5)
Vertical arc (Ml_2)	1	—	1	12	Chinese (4·3), Egyptian: Ptolemaic and Roman (3·9)
100 Ml_2/Ml_1	6	Eskimo (4·0)	0	8	—
Horizontal chord (C)	9	Maori (3·6)	1	4	—
Subtense to chord (S)	5	Negro: Nigeria (3·7)	2	7	—
100 S/C	6	—	2	6	—
	Standard deviations				
Horizontal arc (Ml_1)	8	Egyptian: dynastic (4·1)	0	6	—
Vertical arc (Ml_2)	1	—	0	13	Punjabi (5·1), English (4·8)
100 Ml_2/Ml_1	4	—	0	10	English (3·7)
Horizontal chord (C)	6	—	0	8	—
Subtense to chord (S)	8	—	0	6	Maori (4·2)
100 S/C	8	—	0	6	—

Bilateral comparisons of three constants have been made in Tables III and IV on the assumption that all races tend to show the same asymmetry in type or variability. The validity of this assumption may be examined. In the case of any particular one of the constants, it is clear that in the vast majority of cases the bilateral difference for a series *A* and the corresponding difference for another series *B* will be found to differ insignificantly from one another. By selecting extreme values, however, some significant differences of differences might be found. An examination of Table IV suggests that the clearest evidence of a racial difference in asymmetry, if such exist, is most likely to be shown in the case of the means for the horizontal arc (Ml_1). At one extreme the Hindu series has a mean for the left side which is 0·9 greater than that for the right, and the probable error of this difference is found to be 0·120 on the assumption that the bilateral correlation is the same as that for the long Egyptian series. At the other extreme the Javanese series has a mean for the *right* side which is 0·5 greater than that for the left and the probable error of this difference is found to be 0·144, on the same assumption. The difference of the differences for the two series is 1·4, and this is 7·5 times its probable error (0·187). This appears to afford clear evidence of racial distinction, but it is somewhat uncertain owing to the fact that the Egyptian bilateral correlation may differ appreciably from the unknown Hindu and Javanese values. And it must also be remembered that the case considered is an extreme value in a series of differences, so that a higher ratio of the constant to its probable error must be taken to indicate significance than would be the case if a single difference were being considered by itself. It will be safest to conclude that racial differences in asymmetry are certainly very small, and more abundant material would be needed in order to demonstrate beyond question that any races are differentiated in this way in the case of characters of the malar bone.

5. *The value of the measurements for the purpose of racial classification.* The measurements were taken with the primary object of discovering whether their averages for different series provide suggestive arrangements which might aid attempts to determine the racial relationships of the populations represented. The number of series measured is not large, but it should be sufficient to show which of the characters of the malar bone are likely to be most useful for the purpose in view. The means are given in Table I and the arrangements provided by three different pairs of the characters are shown in Figs. 2–4. In considering these figures it is necessary to appreciate in a general way the differences for each variate which may be taken to indicate clear differentiation.

Fig. 2 shows the inter-racial correlation of the horizontal and vertical arcs, the points being determined by the male means for the left side. In the case of both of these characters most of the differences between the means are large enough to indicate statistical significance. For the Punjabi and Hindu means in the left-hand bottom corner of the diagram, for example, the difference in the

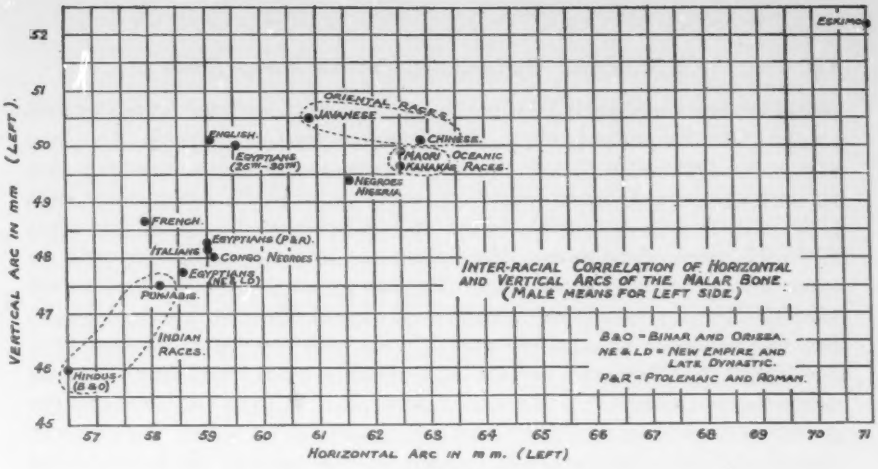


Fig. 2.

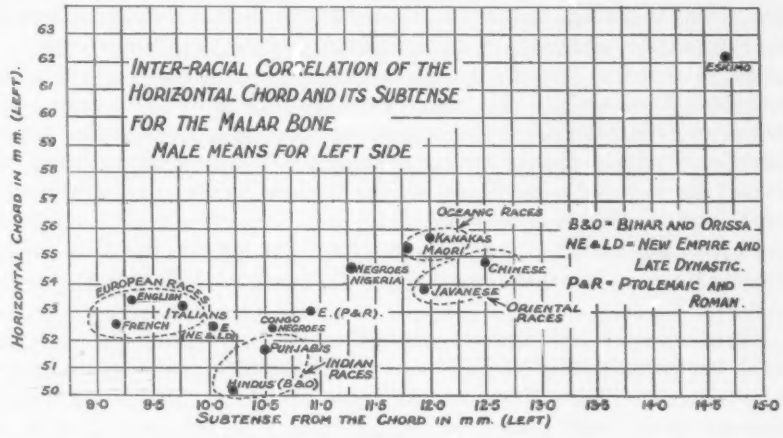


Fig. 3.

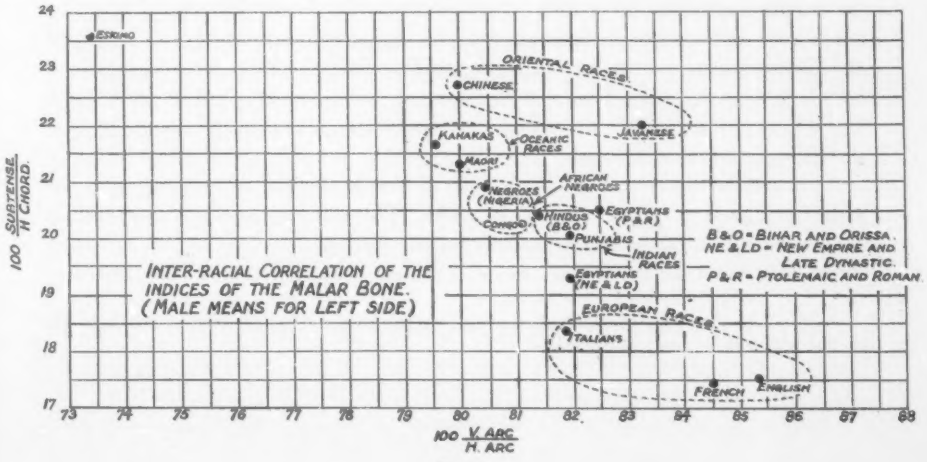


Fig. 4.

case of the horizontal arc is 3·3 times its probable error, and that for the vertical arc 3·8 times its probable error. Both measurements are capable of making many clear distinctions between the racial types, and there is a suggestion that different members of the same family of races show small differences, while most of these families are distinguished from one another by occupying different areas of the bi-variate distribution. More abundant material would obviously be required to substantiate these points. The inter-racial distribution of either arc considered singly appears to be fairly continuous if the Eskimo series is omitted.* This has means widely removed and differing with marked significance from those for all the other series, and the large size of its malar bones appears to be a salient characteristic of the specialized Eskimo type. It should be noted that the measurements are not available for any American-Indian series. The two arcs appear to be highly correlated inter-racially, and the point for the Eskimo series is clearly the one which would be farthest removed from the regression straight line. Owing to the high correlation, we may anticipate that the index expressing the vertical as a percentage of the horizontal arc will be of little interest.

Fig. 3 shows the distribution of the series given by their means for the horizontal chord, and the maximum subtense to this chord. It is again found that most of the differences between the points are statistically significant in the case of each variate, and the Eskimo means again differ from all the others with marked significance. Some of the families of races appear to be separated, as before, and the differences between different races belonging to the same family are apparently all small. The inter-racial correlation between the chord and the subtense seems to be sensible but appreciably lower than that between the two arcs.

Fig. 4 shows the arrangement provided by the two indices. As was anticipated, the index 100 Ml_2/Ml_1 fails to arrange the series in any suggestive order. It is distinguished from all the absolute measurements by the fact that it shows far more insignificant than significant differences in the comparison of all possible pairs of the means, though the Eskimo mean still diverges widely from all others available. The index derived from the horizontal chord and the maximum subtense to it distinguishes the types far more clearly: in this case the Eskimo mean is still extreme, but it differs insignificantly from the Chinese which is nearest to it. This measurement of curvature appears to distinguish the different families of races rather more effectively than any one of the four absolute measurements does.

A provisional estimate of the value of the malar-bone measurements for anthropological purposes can now be given. The material available for them is ample enough to show that racial types of cranium have malar bones which differ very appreciably in both size and shape. Of the four absolute measure-

* The mean of 63·3 for the left horizontal arc given by Dr von Bonin for 50 New Britain skulls is close to the Maori and Kanaka means.

ments and two indices considered, the index obtained by expressing the vertical as a percentage of the horizontal arc is quite the most constant. This appears to be of little value for purposes of racial classification. The other five characters seem to differentiate the racial types as effectively as most of the usual cranial measurements, and more effectively than several of these. The material available suggests that they are characters which tend to be constant for races belonging to the same family of races, but which provide suggestive orders when the different families are compared with one another. They are thus of the same nature as skin colour, the nasal index, measures of prognathism and of the "flatness" of the facial skeleton, and indices derived from the lengths of the limb bones. Stature, the cephalic index and most calvarial measurements differ from these as they fail to make clear distinctions between the different families of races. Among races of the Old World the European and Indian, with their smaller and flatter malar bones, are at one extreme of the range: Oriental races have the largest and most curved bones, and negro and ancient Egyptian occupy intermediate positions. This arrangement accords in a general way with those provided by the indices measuring the "flatness" of the facial skeleton as a whole. In both cases, too, the Eskimo type has been found to occupy markedly aberrant positions. Judging from the short series measured, its malar bones are far larger than those of European, African, Asiatic and Oceanic types; they also show a greater degree of curvature, but the Eskimo type is most clearly distinguished by the fact that the heights of its malar bones (measured by the vertical arcs) are most peculiarly small compared with their antero-posterior lengths (measured by the horizontal arcs). This is of particular interest since the index measuring this ratio is the character least capable of distinguishing the other races from one another. The Eskimo type is detached, as it were, from the continuous system to which the others belong. It is generally recognized to be peculiarly specialized, but none of its characters are known to be more characteristic than these malar-bone measurements and the indices of facial "flatness". More of these data for Eskimo, Eastern Asiatic and American-Indian cranial series would probably throw as much light on the question of the affinities of the Eskimo population as any other new material.

While they are incapable of providing by themselves any reliable classification of the races of modern man, there is every promise that the malar-bone measurements dealt with in this paper will prove to be a valuable aid for the purpose when considered in conjunction with other characters. Hence it is suggested that they might be included with advantage in the routine descriptions of racial series of crania.

THE SAMPLING DISTRIBUTION OF THE CRITERION λ_{H_1} WHEN THE HYPOTHESIS TESTED IS NOT TRUE

[EDITORIAL NOTE. The criterion λ_{H_1} is appropriate to test the statistical hypothesis that the standard deviations of a character x are the same in a number, say k, of different normal populations. In the form $L_1 = \lambda_{H_1}^{2/N}$ (where N is the number of observations in the pooled samples), the criterion becomes the ratio of the weighted geometric to the weighted arithmetic mean of the k sample variances. For the special case where the samples are of equal size, tables of 5 % and 1 % probability levels have been determined by an approximate method by Mr P. P. N. Nayer.*

It is important however not only to have available these significance levels and so to control the risk of rejecting the hypothesis tested when it is true, but also to have some means of appreciating the chance that the test will detect real differences in population standard deviations when they exist. By this means it becomes possible to compare the efficiency of this and alternative tests. Over a year ago Dr S. S. Wilks promptly responded to a request of mine for help in this matter by providing the sampling moments of L_1^{-1}, in the general case where the population standard deviations are unequal. Since then Miss C. M. Thompson has compared his suggested Type III curves, having these moments, with a series of values of L_1^{-1} calculated from experimental sampling data. The correspondence between experiment and the Wilks's curves is excellent. Some further research into the matter is in progress. E.S.P.]

I. NOTE ON THE GENERAL SAMPLING MOMENTS OF λ_{H_1}

BY S. S. WILKS

Neyman & Pearson† have considered in some detail the problem of deriving a criterion λ_{H_1} for testing the hypothesis H_1 that k samples have come from populations with equal variances but with means having any values whatever. They have discussed the sampling theory of λ_{H_1} when H_1 is true. Here we shall be concerned with the more general case in which H_1 is not true. In order to

* *Statistical Research Memoirs* (Dept. of Statistics, University College, London), I (1936), p. 51. The substantial accuracy of the approximation involved has since been verified by Mr U. S. Nair whose work on the subject will be published shortly.

† J. Neyman & E. S. Pearson, "On the Problem of k Samples," *Bull. int. Acad. Cracovie*, Sér. A, 1931, pp. 460–81.

indicate more clearly the point of departure for this note we shall briefly describe what has been done.

Let the tth sample ($t=1, 2, \ldots, k$) of n_t individuals be denoted by Σ_t and suppose Σ_t has been drawn at random from a normal population with mean a_t and variance σ_t^2. Let $\bar{x}_t$ and s_t^2 be the mean and squared standard deviation of Σ_t.* In the present problem $\bar{x}_t$ and s_t^2 ($t=1, 2, \ldots, k$) are the only functions of the individual observations with which we shall be concerned. The probability that $\bar{x}_t$ and s_t^2 will fall in the infinitesimal ranges $\bar{x}_t \pm \frac{1}{2}d\bar{x}_t$, $s_t^2 \pm \frac{1}{2}ds_t^2$ ($t=1, 2, \ldots, k$) will be proportional to

$$C = \prod_{t=1}^{k} \frac{\left(\frac{n_t}{2\sigma_t^2}\right)^{\frac{n_t}{2}} (s_t^2)^{\frac{n_t-3}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{n_t-1}{2}\right)} e^{-\frac{n_t}{2\sigma_t^2}\{s_t^2+(\bar{x}_t-a_t)^2\}} \qquad \ldots\ldots(1)$$

We may regard the k samples $\Sigma_1, \Sigma_2, \ldots, \Sigma_k$ described by the k pairs of quantities $\bar{x}_t$, s_t^2 as having been drawn from the grand population (1).

Now H_1 is the hypothesis that the Σ_t are from populations with

$$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2.$$

The set Ω of admissible populations consists of all populations (1) which could be obtained by taking all possible values of σ_t^2 and a_t. The set ω of populations is that subset of Ω for which the σ's are equal. The criterion λ_{H_1} is the ratio

$$\frac{C(\omega_{\max})}{C(\Omega_{\max})},$$

where $C(\omega_{\max})$ is the maximum of C taken over all populations in ω and $C(\Omega_{\max})$ is the maximum of C taken over all populations in Ω. Expressed in terms of the s's

$$\lambda_{H_1} = \prod_{t=1}^{k} \left(\frac{s_t^2}{s_a^2}\right)^{\frac{n_t}{2}}, \qquad \ldots\ldots(2)$$

where $$Ns_a^2 = \sum_{t=1}^{k} n_t s_t^2, \quad N = n_1 + n_2 + \ldots + n_k.$$

Neyman & Pearson† have considered the sampling properties of λ_{H_1} under the assumption that H_1 is true, that is, that the samples are drawn from a member of ω. We shall consider the sampling properties of λ_{H_1} under the assumption that the samples are from any member of Ω, in which the σ's are not necessarily the same.

Since we are using λ_{H_1} as an instrument for ordering the data embodied in $\Sigma_1, \Sigma_2, \ldots, \Sigma_k$ with respect to the tenability of H_1 it is clearly immaterial whether λ_{H_1} or any single-valued function of λ_{H_1} be used. From a theoretical point of

* Here s_t^2 is defined by $n_t s_t^2 = \sum_{i=1}^{n_t} (x_{it} - \bar{x}_t)^2$.

† *Loc. cit.* pp. 467–73.

view the problem can be somewhat simplified by using $\lambda_{H_1}^{-2/N}$. Since λ_{H_1} is a function of $s_1^2, s_2^2, \ldots, s_k^2$, the gth moment of $\lambda_{H_1}^{-2/N}$ will be defined by the expression

$$\mu_g' = \int_0^\infty \cdots \int_0^\infty \lambda_{H_1}^{-2g/N}\, df_1 \ldots df_k, \qquad \ldots\ldots(3)$$

where

$$df_t = \frac{\left(\frac{n_t}{2\sigma_t^2}\right)^{\frac{n_t-1}{2}} (s_t^2)^{\frac{n_t-3}{2}}}{\Gamma\left(\frac{n_t-1}{2}\right)} e^{-\frac{n_t s_t^2}{2\sigma_t^2}}\, ds_t^2. \qquad \ldots\ldots(4)$$

Now it is clear from (2) that μ_g' will be the gth derivative with respect to θ of the following expression at $\theta = 0$,

$$\phi_g(\theta) = \int_0^\infty \cdots \int_0^\infty (s_1^2)^{-\frac{n_1 g}{N}} \ldots (s_k^2)^{-\frac{n_k g}{N}} e^{s_a{}^2\theta}\, df_1 \ldots df_k$$

$$= K \prod_{t=1}^{k} \left[\frac{\Gamma\left(\frac{n_t-1}{2} - \frac{n_t g}{N}\right)}{\left(\frac{n_t}{2\sigma_t^2} - \frac{n_t}{N}\theta\right)^{\frac{n_t-1}{2} - \frac{n_t g}{N}}} \right], \qquad \ldots\ldots(5)$$

where

$$K = \prod_{t=1}^{k} \frac{\left(\frac{n_t}{2\sigma_t^2}\right)^{\frac{n_t-1}{2}}}{\Gamma\left(\frac{n_t-1}{2}\right)}.$$

For $g = 1, 2$ we find with little difficulty

$$\mu_1' = \prod_{t=1}^{k} \left[\frac{\Gamma\left(\frac{n_t-1}{2} - \frac{n_t}{N}\right)}{\Gamma\left(\frac{n_t-1}{2}\right)} \left(\frac{n_t}{\sigma_t^2}\right)^{\frac{n_t}{N}} \right] \left[\sum_{t=1}^{k} \left(\frac{n_t-1}{2N} - \frac{n_t}{N^2}\right)\sigma_t^2 \right], \qquad \ldots\ldots(6)$$

$$\mu_2' = \prod_{t=1}^{k} \left[\frac{\Gamma\left(\frac{n_t-1}{2} - \frac{2n_t}{N}\right)}{\Gamma\left(\frac{n_t-1}{2}\right)} \left(\frac{n_t}{\sigma_t^2}\right)^{\frac{2n_t}{N}} \right] \left[\left\{ \sum_{t=1}^{k} \left(\frac{n_t-1}{2} - \frac{2n_t}{N}\right) \frac{\sigma_t^2}{N} \right\}^2 \right.$$

$$\left. + \sum_{t=1}^{k} \left(\frac{n_t-1}{2} - \frac{2n_t}{N}\right) \frac{\sigma_t^4}{N^2} \right]. \qquad \ldots\ldots(7)$$

In the important practical case in which $n_1 = n_2 = \ldots = n_k = n$, (6) and (7) reduce to

$$\mu_1' = \left[\frac{\Gamma\left(\frac{n-1}{2} - \frac{1}{k}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \right]^k \left(\frac{n-1}{2k} - \frac{1}{k^2}\right) \left(\frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_k^2)}{(\sigma_1^2 \sigma_2^2 \ldots \sigma_k^2)^{\frac{1}{k}}} \right), \qquad \ldots\ldots(8)$$

$$\mu_2' = \left[\frac{\Gamma\left(\frac{n-1}{2}-\frac{2}{k}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right]^k \left[\left(\frac{n-1}{2}-\frac{2}{k}\right)^2 \left(\frac{\sigma_1^2+\sigma_2^2+\ldots+\sigma_k^2}{k(\sigma_1^2\sigma_2^2\ldots\sigma_k^2)^{\frac{1}{k}}}\right)^2 + \frac{\left(\frac{n-1}{2}-\frac{2}{k}\right)(\sigma_1^4+\sigma_2^4+\ldots+\sigma_k^4)}{k^2(\sigma_1^2\sigma_2^2\ldots\sigma_k^2)^{\frac{2}{k}}}\right]. \quad \ldots\ldots(9)$$

When the hypothesis H_1 is true, that is, when $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$, it can be easily verified that

$$\mu_g' = \frac{\Gamma\left(\frac{N-k}{2}\right)}{\Gamma\left(\frac{N-k}{2}-g\right)} \prod_{l=1}^{k} \left[\left(\frac{n_l}{N}\right)^{\frac{n_l g}{N}} \frac{\Gamma\left(\frac{n_l-1}{2}-\frac{n_l g}{N}\right)}{\Gamma\left(\frac{n_l-1}{2}\right)}\right], \quad \ldots\ldots(10)$$

as obtained by Neyman & Pearson. μ_g' will exist for all values of g for which the arguments of the gamma functions are positive.

The higher moments of $\lambda_{H_1}^{-2/N}$ become more and more complicated so that there is little hope of finding a workable form of the exact distribution function of $\lambda_{H_1}^{-2/N}$. Therefore, since $\lambda_{H_1}^{-2/N}$ has the range 1 to ∞, it appears that its distribution could be reasonably approximated by fitting a Type III curve by means of the first two moments. Let the form of the curve be

$$\frac{e^b b^a}{\Gamma(a)}(x-1)^{a-1}e^{-bx}. \quad \ldots\ldots(11)$$

Equating the first two moments of (11) about the origin to μ_1' and μ_2' and solving for a and b, we find

$$\left.\begin{aligned} a &= \frac{(\mu_1'-1)^2}{\mu_2'-\mu_1'^2} \\ b &= \frac{\mu_1'-1}{\mu_2'-\mu_1'^2} \end{aligned}\right\}. \quad \ldots\ldots(12)$$

II. AN INVESTIGATION INTO THE ADEQUACY OF DR WILKS'S CURVES

BY CATHERINE M. THOMPSON

The basic data used consisted of 500 samples of (i) $n_1 = 5$, (ii) $n_2 = 10$ and (iii) $n_3 = 15$ from a common normal population, obtained with the help of Tippett's Random Numbers. The values of, say, $v = \Sigma(x-\bar{x})^2/n$ had already been calculated for each of the 1500 samples for another purpose. By multiplying the values of v by appropriate factors it was possible to obtain 500 sets of values s_1^2, s_2^2 and s_3^2 from populations having unequal variances σ_1^2, σ_2^2 and σ_3^2, respectively.

For example, if v_1, v_2 and v_3 denote the basic sample variances, if the common population standard deviation is unity and if $N = n_1+n_2+n_3 = 30$, then

$$L_1 = \frac{\prod_{i=1}^{3} (s_i^2)^{\frac{n_i}{N}}}{\frac{1}{N}\sum_{i=1}^{3}(n_i s_i^2)} = \frac{v_1^{\frac{1}{6}} v_2^{\frac{2}{3}} v_3^{\frac{1}{6}} \times (\sigma_1^2)^{\frac{1}{6}} (\sigma_2^2)^{\frac{2}{3}} (\sigma_3^2)^{\frac{1}{6}}}{\frac{1}{6}\sigma_1^2 v_1 + \frac{2}{3}\sigma_2^2 v_2 + \frac{1}{6}\sigma_3^2 v_3}. \qquad \text{.....(1)}$$

The result depends only on the relative magnitude of the three values of σ^2. Six different cases were taken and for each the 500 values of L_1^{-1} were computed; since the same basic values of v_1, v_2 and v_3 were used, the six resulting frequency distributions of L_1^{-1} are not completely independent, but their relationship is of no simple character. The cases taken were as follows:

	σ_1^2 :	σ_2^2 :	σ_3^2
Case 1	2	1	2
,, 2	1	2	2
,, 3	2	1	1
,, 4	1	2	1
,, 5	4	2	1
,, 6	10	2	1

Since the three values of n are unequal, the first four cases correspond to different situations. The resulting histograms for the six distributions of L_1^{-1} are shown in Figs. 1 and 2.

For each case the following steps were then performed:

(i) The appropriate moments μ_1' and μ_2' and hence $\mu_2 = \mu_2' - (\mu_1')^2$ were calculated from Wilks's equations (6) and (7).

(ii) These moment values were inserted into his equation (12) to give the constants a and b.

(iii) These constants were inserted in turn into his Type III equation (11), the curves drawn and frequencies calculated to enable a χ^2 test to be applied.

A summary of results is shown in Table I. The column headed $P\{\chi^2 > \chi_0^2\}$ shows the result of applying the χ^2 test for goodness of fit, χ_0^2 being the observed value. The agreement between the theoretical curves and the experimental sampling results is very close, and suggests that the method of approximation to the unknown true distribution of L_1^{-1} is most satisfactory for practical purposes.

Of course the investigation only covers the case $k = 3$ and $n_1 = 5$, $n_2 = 10$, $n_3 = 15$, but at any rate for larger samples one would not expect worse agreement.

It is also of interest to investigate, in the six cases, what would be the chance of detecting from the samples that σ_1, σ_2 and σ_3 were not all equal. Suppose that we used a rule of rejecting the hypothesis, H_0, that $\sigma_1 = \sigma_2 = \sigma_3$ whenever L_1

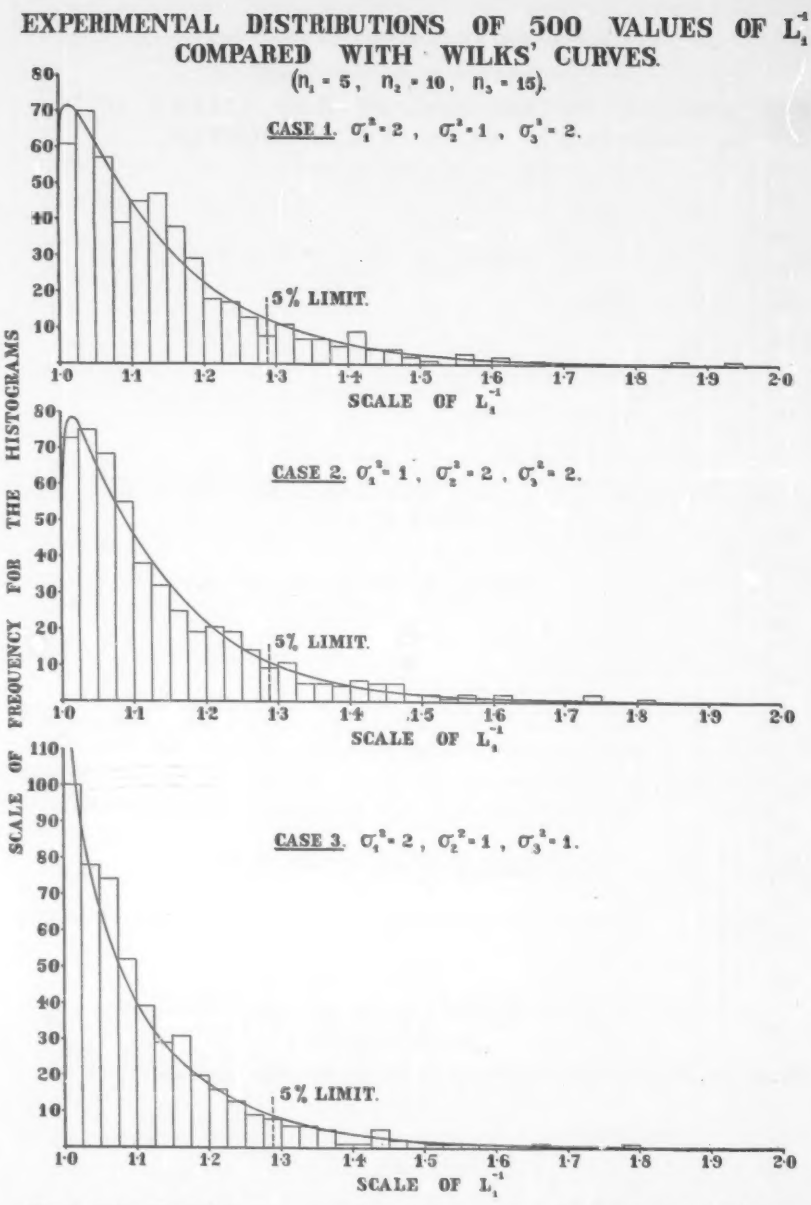

EXPERIMENTAL DISTRIBUTIONS OF 500 VALUES OF L_1^{-1} COMPARED WITH WILKS' CURVES.
($n_1 = 5$, $n_2 = 10$, $n_3 = 15$).
CASE 1. $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $\sigma_3^2 = 2$.
5% LIMIT.
SCALE OF L_1^{-1}
CASE 2. $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\sigma_3^2 = 2$.
5% LIMIT.
SCALE OF L_1^{-1}
CASE 3. $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $\sigma_3^2 = 1$.
5% LIMIT.
SCALE OF L_1^{-1}
SCALE OF FREQUENCY FOR THE HISTOGRAMS

Fig. 1. N.B. The 5 % limit is appropriate for the case where the hypothesis tested is true.

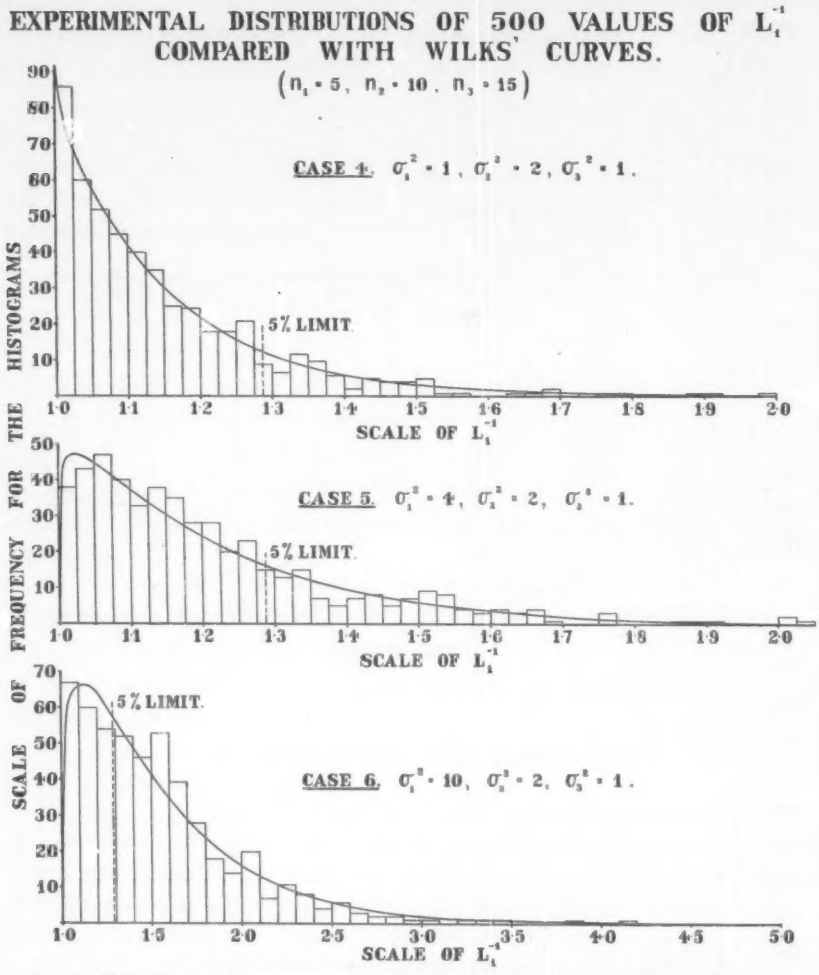

EXPERIMENTAL DISTRIBUTIONS OF 500 VALUES OF L_1^{-1} COMPARED WITH WILKS' CURVES.
$(n_1 = 5,\ n_2 = 10,\ n_3 = 15)$
CASE 4. $\sigma_1^2 = 1,\ \sigma_2^2 = 2,\ \sigma_3^2 = 1.$
CASE 5. $\sigma_1^2 = 4,\ \sigma_2^2 = 2,\ \sigma_3^2 = 1.$
CASE 6. $\sigma_1^2 = 10,\ \sigma_2^2 = 2,\ \sigma_3^2 = 1.$
5% LIMIT.
SCALE OF L_1^{-1}
SCALE OF FREQUENCY FOR THE HISTOGRAMS

Fig. 2. N.B. The 5 % limit is appropriate for the case where the hypothesis tested is true.

TABLE I

Summary of results of calculations

Case	Values of the σ^2's	Moments	Theory	Observed	a	b	$P\{\chi^2 > \chi_0^2\}$	Power of test*
1	2 : 1 : 2	μ_1'	1·1462	1·1426	1·0891	7·4493	·615	·136
		μ_2	0·0196	0·0168				
2	1 : 2 : 2	μ_1'	1·1356	1·1325	1·0807	7·9725	·965	·120
		μ_2	0·0170	0·0160				
3	2 : 1 : 1	μ_1'	1·1152	1·1082	0·8197	7·1144	·525	·093
		μ_2	0·0162	0·0115				
4	1 : 2 : 1	μ_1'	1·1553	1·1495	0·9747	6·2760	·853	·158
		μ_2	0·0247	0·0230				
5	4 : 2 : 1	μ_1'	1·2195	1·2071	1·1913	5·0165	·816	·271
		μ_2	0·0438	0·0330				
6	10 : 2 : 1	μ_1'	1·5663	1·5283	1·2543	2·2148	·389	·646
		μ_2	0·2557	0·1968				

* Using 5 % significance level.

falls beyond the 5 % level, say $L_1(0{\cdot}05)$. It is first necessary to calculate this level for the particular case considered. To do this, Neyman & Pearson's Type I approximation* to the distribution of L_1 if H_0 is true, namely

$$p(L_1) = \frac{\Gamma(m_1+m_2)}{\Gamma(m_1)\,\Gamma(m_2)}\, L_1^{m_1-1}(1-L_1)^{m_2-1} \qquad \text{.....(2)}$$

may be used. The true sampling moments of L_1 about zero are in this case

$$\left.\begin{aligned} \mu_1' &= \frac{2N}{N-3}\prod_{l=1}^{k}\left\{\frac{\Gamma\left(\frac{n_l-1}{2}+\frac{n_l}{N}\right)}{n_l^{\frac{n_l}{N}}\,\Gamma\left(\frac{n_l-1}{2}\right)}\right\} = 0{\cdot}9225\,2768, \\ \mu_2' &= \frac{4N^2}{(N-3)(N-1)}\prod_{l=1}^{3}\left\{\frac{\Gamma\left(\frac{n_l-1}{2}+\frac{2n_l}{N}\right)}{n_l^{\frac{2n_l}{N}}\,\Gamma\left(\frac{n_l-1}{2}\right)}\right\} = 0{\cdot}8561\,9647. \end{aligned}\right\} \qquad \text{.....(3)}$$

* See reference on p. 124 above and also P. P. N. Nayer, *loc. cit.*

The values for m_1 and m_2 are then chosen so that the first two moments of the distribution (2) have the values (3), that is to say,

$$m_1 = \frac{\mu_1'(\mu_1'-\mu_2')}{\mu_2'-(\mu_1')^2} = 11{\cdot}90710, \quad m_2 = \frac{(1-\mu_1')(\mu_1'-\mu_2')}{\mu_2'-(\mu_1')^2} = 0{\cdot}99994. \quad \ldots\ldots(4)$$

In this case the value of m_2 is so close to unity that we may use the approximation

$$p(L_1) = m_1 L_1^{m_1-1}, \quad \ldots\ldots(5)$$

and hence for the 5 % limit

$$0{\cdot}05 = \int_0^{L_1(0{\cdot}05)} p(L_1)\,dL_1 = \{L_1(0{\cdot}05)\}^{m_1}$$

and

$$L_1(0{\cdot}05) = 0{\cdot}778. \quad \ldots\ldots(6)$$

This limit, or rather the limit $\{L_1(0{\cdot}05)\}^{-1} = 1{\cdot}286$, has been drawn in each of the diagrams. Neyman & Pearson* have defined the power of a test with regard to an alternative hypothesis H_1 as the probability that it will reject the hypothesis tested, H_0, when H_1 is true. Thus if a 5 % level of significance is used, the power of the L_1 test in the six cases illustrated is given by the proportionate area under Wilks's curves lying to the right of the critical levels drawn at $L_1^{-1} = 1{\cdot}286$. The six values of this probability, obtained by quadrature, are given in Table I.

It is interesting to note how, owing to the three samples being of unequal size, the power is different in each of the first four cases. Thus when the samples of 10 and 15 come from populations with the same variance, $\sigma_2^2 = \sigma_3^2$, and the smallest sample of 5 from a population with twice the variance (case 3), the test is least likely to detect the difference. It is somewhat more likely to do so when $\sigma_1^2 = \frac{1}{2}\sigma_2^2 = \frac{1}{2}\sigma_3^2$ (case 2). We also see how difficult it is to detect differences in population variances when only small samples are available; even in case 6, where $\sigma_1^2 : \sigma_2^2 : \sigma_3^2 = 10 : 2 : 1$, the odds are only 2 to 1 in favour of our being able to discover the difference, using the L_1 test.

* Neyman & Pearson, "Contribution to the Theory of testing Statistical Hypotheses," *Statistical Research Memoirs*, I (1936), pp. 1–37.

THE EXACT VALUE OF THE MOMENTS OF THE DISTRIBUTION OF χ^2, USED AS A TEST OF GOODNESS OF FIT, WHEN EXPECTATIONS ARE SMALL

By J. B. S. HALDANE, F.R.S.

1. Introduction to general method

In genetical practice we are constantly presented with large numbers of small samples from populations consisting of several well-defined classes. For example in the mouse we can readily obtain hundreds of litters containing anything from one up to about twelve members. Their totals may agree satisfactorily with expectation on a Mendelian basis, for example $\frac{1}{2}$ coloured, $\frac{1}{2}$ white, or $\frac{9}{16}$ grey, $\frac{3}{16}$ black, $\frac{1}{4}$ white. But we desire to know whether the individual litters can be regarded as random samples from such a population. In addition the problem of homogeneity may arise. That is to say the population as a whole may not conform to any particular expectation. But we may desire to know whether the litters can be regarded as random samples of the population given by the totals.

It has long been known that when the numbers expected in any observation are small, the distribution of χ^2 departs from that given by Pearson (1900). The mean appears sometimes, but not always, to be equal to the number of degrees of freedom. But the variance is no longer exactly equal to twice that number. Exact expressions for it in certain cases have been given by Pearson (1932) and Cochran (1936). These are based on an ingenious application of the theory of multiple contingency by Pearson.

It will be shown in this paper that the first few moments can often be calculated by entirely elementary methods involving nothing more advanced than the multinomial theorem. In an accompanying paper (Grüneberg and Haldane, 1937) they will be applied to actual data on mice.

We first study the distribution of χ^2 in a n-fold table with $n-1$ degrees of freedom, then in a $(m \times n)$-fold table with $m(n-1)$ degrees of freedom. For genetical work we are particularly interested in the $(n \times 2)$-fold table with n degrees of freedom. As a limiting case of the 2-fold table with 1 degree of freedom we derive the moments of the variance of samples from a Poisson series, and thence the distribution of χ^2 in a n-fold table with n degrees of freedom. The important case of the $(m \times n)$-fold table with $(m-1)(n-1)$ degrees of freedom remains to be investigated.

Consider a sample of s individuals falling into n classes. Let the expected and observed numbers in these classes be:

$$\text{Expected } p_1 s, p_2 s, \ldots, p_i s, \ldots, p_n s,$$

$$\text{Observed } a_1, a_2, \ldots, a_i, \ldots, a_n,$$

where $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} a_i = s$. It is assumed that we are sampling from an infinite population, or that if it is finite, any individual observed is replaced before the next individual is chosen at random. We shall use the notation $s_n = \dfrac{s!}{(s-n)!}$, and $E[x]$ to denote the expected value of x, or $\bar{x}$.

The probability of obtaining in a sample exactly a_1, a_2, a_3, etc., members of the n classes is $s! \prod_{i=1}^{n} \dfrac{p_i^{a_i}}{a_i!}$.

Hence the expected value of

$$\frac{a_f!}{(a_f-\alpha_f)!}\frac{a_g!}{(a_g-\alpha_g)!}\frac{a_h!}{(a_h-\alpha_h)!}$$

is the sum of this quantity multiplied by

$$s! \prod_{i=1}^{n} \frac{p_i^{a_i}}{a_i!},$$

summation being over all permissible sets of values of a_f, a_g and a_h, i.e. for all zero or positive integral values satisfying the condition $\sum_{i=1}^{n} (a_i) = s$.* Making use of the multinomial theorem it is seen that this sum is $p_f^{\alpha_f} p_g^{\alpha_g} p_h^{\alpha_h} s_{(\alpha_f+\alpha_g+\alpha_h)}$. But we can readily express any power of a_f, or any multiple of powers such as $a_f^4 a_g^2 a_h^2$, as a sum of expressions of the form $\dfrac{a_i!}{(a_i-\alpha_i)!}$ and of their products. Hence we can express the expected value of any power or product of powers as a sum of terms of the form $p_f^r p_g^s p_h^t s_{r+s+t}$. The following expressions for the expected values of powers and their products will be required in the analysis which follows:

$$E[a_i^2] = p_i^2 s_2 + p_i s,$$

$$E[a_i^4] = p_i^4 s_4 + 6p_i^3 s_3 + 7p_i^2 s_2 + p_i s,$$

$$E[a_i^6] = p_i^6 s_6 + 15p_i^5 s_5 + 65p_i^4 s_4 + 90p_i^3 s_3 + 31p_i^2 s_2 + p_i s,$$

$$E[a_i^8] = p_i^8 s_8 + 28p_i^7 s_7 + 266p_i^6 s_6 + 1050p_i^5 s_5 + 1701p_i^4 s_4 + 966p_i^3 s_3 + 127p_i^2 s_2 + p_i s.$$

In general $E[a_i^n] = \Sigma_r U_n p_i^r s_r$, where ${}_rU_n = {}_{r-1}U_{n-1} + r\,{}_rU_{n-1}$.†

$$E[a_i^2 a_j^2] = p_i^2 p_j^2 s_4 + (p_i^2 p_j + p_i p_j^2) s_3 + p_i p_j s_2,$$

$$E[a_i^4 a_j^2] = p_i^4 p_j^2 s_6 + (p_i^4 p_j + 6p_i^3 p_j^2) s_5 + (6p_i^3 p_j + 7p_i^2 p_j^2) s_4 + (7p_i^2 p_j + p_i p_j^2) s_3 + p_i p_j s_2,$$

$$E[a_i^2 a_j^2 a_k^2] = p_i^2 p_j^2 p_k^2 s_6 + (p_i p_j^2 p_k^2 + p_i^2 p_j p_k^2 + p_i^2 p_j^2 p_k) s_5 + (p_i^2 p_j p_k + p_i p_j^2 p_k + p_i p_j p_k^2) s_4 + p_i p_j p_k s_3,$$

* The value of $a_i!/(a_i-\alpha_i)! = a_i(a_i-1)\ldots(a_i-\alpha_i+1)$ is of course zero for $a_i < \alpha_i$.

† Mr C. Eisenhart has kindly pointed out to me that these coefficients are differences of powers of zero divided by appropriate factorials.

$$E[a_i^6 a_j^2] = p_i^6 p_j^2 s_8 + (p_i^6 p_j + 15 p_i^5 p_j^2) s_7 + (15 p_i^5 p_j + 65 p_i^4 p_j^2) s_6 + (65 p_i^4 p_j + 90 p_i^3 p_j^2) s_5 + (90 p_i^3 p_j + 31 p_i^2 p_j^2) s_4 + (31 p_i^2 p_j + p_i p_j^2) s_3 + p_i p_j s_2,$$

$$E[a_i^4 a_j^4] = p_i^4 p_j^4 s_8 + (6 p_i^4 p_j^3 + 6 p_i^3 p_j^4) s_7 + (7 p_i^4 p_j^2 + 36 p_i^3 p_j^3 + 7 p_i^2 p_j^4) s_6 + (p_i^4 p_j + 42 p_i^3 p_j^3 + 42 p_i^2 p_j^3 + p_i p_j^4) s_5 + (6 p_i^3 p_j + 49 p_i^2 p_j^2 + 6 p_i p_j^3) s_4 + (7 p_i^2 p_j + 7 p_i p_j^2) s_3 + p_i p_j s_2,$$

$$E[a_i^4 a_j^2 a_k^2] = p_i^4 p_j^2 p_k^2 s_8 + (p_i^4 p_j^2 p_k + p_i^4 p_j p_k^2 + 6 p_i^3 p_j^2 p_k^2) s_7 + (p_i^4 p_j p_k + 6 p_i^3 p_j^2 p_k + 6 p_i^3 p_j p_k^2 + 7 p_i^2 p_j^2 p_k^2) s_6 + (6 p_i^3 p_j p_k s_5 + 7 p_i^2 p_j^2 p_k + 7 p_i^2 p_j p_k^2 + p_i p_j^2 p_k^2) s_5 + (7 p_i^2 p_j p_k + p_i p_j^2 p_k + p_i^2 p_j p_k^2) s_4 + p_i p_j p_k s_3,$$

$$E[a_i^2 a_j^2 a_k^2 a_l^2] = p_i^2 p_j^2 p_k^2 p_l^2 s_8 + \Sigma p_i^2 p_j^2 p_k^2 p_l s_7 + \Sigma p_i^2 p_j^2 p_k p_l s_6 + \Sigma p_i^2 p_j p_k p_l s_5 + p_i p_j p_k p_l s_4.$$

In the special case where $p_1 = p_2 = p_3 = \ldots = n^{-1}$ we have such expressions as

$$E[a_i^4 a_j^2 a_k^2] = n^{-8} s_8 + 8 n^{-7} s_7 + 20 n^{-6} s_6 + 21 n^{-5} s_5 + 9 n^{-4} s_4 + s_3.$$

In what follows, Σ denotes summation over all values of i, $\Sigma\Sigma$ summation over all pairs of unequal values of i and j. The following notation is used for sums of reciprocals: $R_1 = \Sigma p_i^{-1}$, $R_2 = \Sigma p_i^{-2}$, $R_3 = \Sigma p_i^{-3}$.

$$\chi^2 = \Sigma \frac{(p_i s - a_i)^2}{p_i s}$$

$$= \frac{1}{s} \Sigma \frac{a_i^2}{p_i} - s.$$

Hence
$$\overline{\chi^2} = s^{-1} (s_2 \Sigma p_i + s \Sigma 1) - s$$
$$= n - 1.$$

$$\overline{\chi^4} = E\left[s^{-2} \Sigma \frac{a_i^4}{p_i^2} + 2 s^{-2} \Sigma\Sigma \frac{a_i^2 a_j^2}{p_i p_j} - 2 \Sigma \frac{a_i^2}{p_i} + s^2 \right]$$

$$= s^{-2} [s_4 (\Sigma p_i^2 + 2 \Sigma\Sigma p_i p_j) + s_3 (6 \Sigma p_i + 4 \Sigma\Sigma p_i) + s_2 (7 \Sigma_1 + 2 \Sigma\Sigma 1) + s \Sigma p_i^{-1}] - 2 (s_2 \Sigma p_i + s \Sigma 1) + s^2$$

$$= s^{-2} [s_4 + (2n + 4) s_3 + (n^2 + 6n) s_2 + R_1 s] - 2 (s_2 + ns) + s^2$$

$$= n^2 - 1 + (R_1 - n^2 - 2n + 2) s^{-1}.$$

Hence
$$\mu_2 = 2 (n - 1) + (R_1 - n^2 - 2n + 2) s^{-1}.$$

This agrees with Pearson's (1932) result. The calculation of the higher moments is somewhat tedious. It can be greatly simplified by the following device. The moments are calculated for the special case when all p_i's are equal. The terms in the general case which involve sums of negative powers of the p_i's are then calculated separately, and an adjustment made to the previous formulae, since when $p_i = n^{-1}$, $R_k = n^{k+1}$.

2. CASE OF AN n-FOLD CLASSIFICATION ($n-1$ DEGREES OF FREEDOM)

For a sample of s divided into n classes of which the expectations are equal, we find:

$$E[\Sigma a_i^2] = n^{-1}(s_2 + ns) = n^{-1}[s^2 + (n-1)s],$$

$$E[\Sigma a_i^2]^2 = n^{-2}[s_4 + (2n+4)s_3 + (n^2+6n)s_2 + n^2 s]$$
$$= n^{-2}[s^4 + 2(n-1)s^3 + (n^2-1)s^2 - 2(n-1)s],$$

$$E[\Sigma a_i^2]^3 = n^{-3}[s_6 + (3n+12)s_5 + (3n^2+30n+32)s_4 + (n^3+21n^2+68n)s_3 + (3n^3+28n^2)s_2 + n^3 s]$$
$$= n^{-3}[s^6 + 3(n-1)s^5 + 3(n^2-1)s^4 + (n^3+3n^2-7n+3)s^3 - 2(n^2+12n-13)s^2 - 4(n^2-7n+6)s],$$

$$E[\Sigma a_i^2]^4 = n^{-4}[s_8 + (4n+24)s_7 + (6n^2+84n+176)s_6 + (4n^3+102n^2 - 544n + 400)s_5 + (n^4+48n^3+516n^2+1136n)s_4 + (6n^4+152n^3+808n^2)s_3 + (7n^4+120n^3)s_2 + n^4 s]$$
$$= n^{-4}[s^8 + 4(n-1)s^7 + 6(n^2-1)s^6 + 4(n^3+3n^2-4n)s^5 + (n^4+8n^3+6n^2-104n+89)s^4 + 4(n^3-17n^2-45n+61)s^3 - 4(2n^3+51n^2-314n+261)s^2 - 8(n^3-31n^2+120n-90)s].$$

Hence

$$\overline{\chi^2} = n-1,$$

$$\overline{\chi^4} = n^2 - 1 - 2(n-1)s^{-1},$$

$$\overline{\chi^6} = (n+3)(n+1)(n-1) - 2(n-1)(n+13)s^{-1} - 4(n-1)(n-6)s^{-2},$$

$$\overline{\chi^8} = (n+5)(n+3)(n+1)(n-1) + 4(n-1)(n^2-12n-85)s^{-1} - 4(n-1)(2n^2+53n-261)s^{-2} - 8(n-1)(n^2-30n+90)s^{-3}.$$

The first four moments and cumulants are:

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = n-1,\\
&\mu_2 = \kappa_2 = 2(n-1) - 2(n-1)s^{-1} = 2s^{-1}(n-1)(s-1),\\
&\mu_3 = \kappa_3 = 8(n-1) + 4(n-1)(n-8)s^{-1} - 4(n-1)(n-6)s^{-2}\\
&\qquad = 4s^{-2}(n-1)(s-1)(n+2s-6),\\
&\mu_4 = 12(n-1)(n+3) + 24(n-1)(3n-19)s^{-1} + 4(n-1)(2n^2-81n+285)s^{-2}\\
&\qquad - 8(n-1)(n^2-30n+90)s^{-3},\\
&\kappa_4 = 48(n-1) + 96(n-1)(n-5)s^{-1} + 8(n-1)(n^2-42n+144)s^{-2}\\
&\qquad - 8(n-1)(n^2-30n+90)s^{-3}\\
&\quad = 8s^{-2}(n-1)(s-1)[n^2 + 6(2s-5)n + 6(s^2-9s+15)],
\end{aligned}\right\} \quad \ldots\ldots(1)$$

where $\kappa_4 = \mu_4 - 3\mu_2^2$.

Hence $$\beta_1=\frac{2(n+2s-6)^2}{(n-1)s(s-1)}.$$

In the general case where not all the p_i's are equal, we find:

$$E\left[\Sigma\frac{a_i^2}{p_i}\right]=s_2+ns=s^2+(n-1)s,$$

$$E\left[\Sigma\frac{a_i^2}{p_i}\right]^2=s_4+(2n+4)s_3+(n^2+6n)s_2+R_1s$$
$$=s^4+2(n-1)s^3+(n^2-1)s^2+(R_1-n^2-2n+2)s,$$

$$E\left[\Sigma\frac{a_i^2}{p_i}\right]^3=s_6+(3n+12)s_5+(3n^2+30n+32)s_4+(3R_1+n^3+18n^2+68n)s_3$$
$$+(3n+28)R_1s_2+R_1s$$
$$=s^6+3(n-1)s^5+3(n^2-1)s^4+(3R_1+n^3-7n+3)s^3$$
$$+[(3n+19)R_1-3n^3-21n^2-24n+26]s^2$$
$$+[R_2-(3n+22)R_1+2n^3+18n^2+28n-24]s,$$

$$E\left[\Sigma\frac{a_i^2}{p_i}\right]^4=s_8+(4n+24)s_7+(6n^2+84n+176)s_6+(6R_1+4n^3+96n^2+544n+400)s_5$$
$$+[(12n+36)R_1+n^4+36n^3+380n^2+1136n]s_4$$
$$+[4R_2+(6n^2+148n+808)R_1]s_3+[(4n+120)R_2+3R_1^2]s_2+R_3s$$
$$=s^8+4(n-1)s^7+6(n^2-1)s^6+(6R_1+4n^3+6n^2-16n)s^5$$
$$+[4(3n+19)R_1+n^4-4n^3-70n^2-104n+89]s^4$$
$$+[4R_2+(6n^2+76n+202)R_1-6n^4-76n^3-270n^2-180n+244]s^3$$
$$+[(4n+108)R_2+3R_1^2-(18n^2+312n+1228)R_1+11n^4+196n^3$$
$$+1024n^2+1256n-1044]s^2+[R_3-(4n+112)R_2-3R_1^2$$
$$+(12n^2+224n+944)R_1-6n^4-120n^3-696n^2-960n+720]s.$$

But $$\chi^{2n}=\left(s^{-1}\Sigma\frac{a_i^2}{p_i}-s\right)^n.$$

Hence

$$\overline{\chi^2}=n-1,$$

$$\overline{\chi^4}=(n+1)(n-1)+(R_1-n^2-2n+2)s^{-1},$$

$$\overline{\chi^6}=(n+3)(n+1)(n-1)+[(3n+19)R_1-(3n^3+21n^2+24n-26)]s^{-1}$$
$$+[R_2-(3n+22)R_1+2n^3+18n^2+28n-24]s^{-2},$$

$$\overline{\chi^8}=(n+5)(n+3)(n+1)(n-1)+2[(3n^2+44n+145)R_1-(32n^4+42n^3+171n^2$$
$$+146n-170)]s^{-1}+[4(n+27)R_2+3R_1^2-2(9n^2+156n+614)R_1$$
$$+11n^4+196n^3+1024n^2+1256n-1044]s^{-2}+[R_3-4(n+28)R_2-3R^2$$
$$+4(3n^2+56n+236)R_1-6(n^4+20n^3+116n^2+160n-120)]s^{-3}.$$

Therefore:

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = n-1,\\
&\mu_2 = \kappa_2 = 2(n-1) + [R_1 - (n^2+2n-2)]\,s^{-1},\\
&\mu_3 = \kappa_3 = 8(n-1) + 2[11R_1 - (9n^2+18n-16)]\,s^{-1}\\
&\qquad\qquad + [R_2 - (3n+22)R_1 + 2(n^3+9n^2+14n-12)]\,s^{-2},\\
&\mu_4 = 12(n-1)(n+3) + 12[(n+31)R_1 - (n^3+25n^2+44n-38)]\,s^{-1}\\
&\quad + [112R_2 + 3R_1^2 - 2(3n^2+118n+658)R_1 + 3(n^4+44n^3+328n^2+488n-380)]\,s^{-2}\\
&\quad + [R_3 - 4(n+28)R_2 - 3R_1^2 + 4(3n^2+56n+236)R_1\\
&\qquad\qquad - 6(n^4+20n^3+116n^2+160n-120)]\,s^{-3},\\
&\kappa_4 = 48(n-1) + 96(4R_1 - 3n^2 - 6n + 5)\,s^{-1}\\
&\quad + 8[14R_2 - 2(14n+83)R_1 + 3(5n^3+41n^2+62n-48)]\,s^{-2}\\
&\quad + [R_3 - 4(n+28)R_2 - 3R_1^2 + 4(3n^2+56n+236)R_1\\
&\qquad\qquad - 6(n^4+20n^3+116n^2+160n-120)]\,s^{-3}.
\end{aligned}\right\} \quad \ldots\ldots(2)$$

It will be seen that when any of the expected frequencies p_i is very small, the moments may be considerably larger than those of the classical χ^2, to which they approximate when the number s in the sample is large.

3. Special case of two classes (1 degree of freedom)

When there are only two expected classes, with frequencies p and q, i.e. $n=2$, we have, if $k = \dfrac{1}{pq}$,

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = 1,\\
&\mu_2 = \kappa_2 = 2 + (k-6)\,s^{-1},\\
&\mu_3 = \kappa_3 = 8 + 2(11k-6)\,s^{-1} + (k^2-30k+120)\,s^{-2},\\
&\mu_4 = 60 + 12(33k-158)\,s^{-1} + (115k^2-2036k+6828)\,s^{-2}\\
&\qquad\qquad + (k^3-126k^2+1680k-5040)\,s^{-3},\\
&\kappa_4 = 48 + 96(4k-19)\,s^{-1} + 16(7k^2-125k+420)\,s^{-2}\\
&\qquad\qquad + (k^3-126k^2+1680k-5040)\,s^{-3}.
\end{aligned}\right\} \quad \ldots\ldots(3)$$

These expressions can of course be calculated independently, and furnish a useful check on the equations (2). When s tends to infinity, provided neither p nor q tends to zero, we have $\kappa_2=2$, $\kappa_3=8$, $\kappa_4=48$, $\kappa_n = 2^{n-1}(n-1)!$, the values appropriate to Pearson's χ^2. If, however, sp remains equal to g while s tends to infinity we have:

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = 1,\\
&\mu_2 = \kappa_2 = 2 + g^{-1},\\
&\mu_3 = \kappa_3 = 8 + 22g^{-1} + g^{-2},\\
&\mu_4 = 60 + 396g^{-1} + 115g^{-2} + g^{-3},\\
&\kappa_4 = 48 + 384g^{-1} + 112g^{-2} + g^{-3}.
\end{aligned}\right\} \quad \ldots\ldots(4)$$

These are the moments and cumulants of $\chi^2=(a-g)^2/g$ for a sample from a Poisson series when the expected value is g and the observed value is a. If V be the variance, $V=(a-g)^2=g\chi^2$, so $\mu_n g^n$ and $\kappa_n g^n$ are the moments and cumulants of the variance of such a sample.

4. Case of $(m\times n)$-fold classification ($m(n-1)$ degrees of freedom)

We now consider the values of χ^2 in 2 dimensional tables. Consider a $(m\times n)$-fold table where m samples of $s_1, s_2, s_3, \ldots, s_r, \ldots, s_m$ members have been drawn independently from an infinite population in which the frequencies of n classes are $p_1, p_2, p_3, \ldots, p_i, \ldots, p_n$, and where, as above,

$$R_1=\sum_{i=1}^{n} p_i^{-1}, \quad R_2=\sum_{i=1}^{n} p_i^{-2}, \quad R_3=\sum_{i=1}^{n} p_i^{-3}.$$

There are clearly $m(n-1)$ degrees of freedom. Summing the cumulants, given in equation (2), appropriate to the χ^2 calculated from each sample, we have therefore:

$$\left.\begin{aligned}
&\mu_1'=\kappa_1=m(n-1),\\
&\mu_2=\kappa_2=2m(n-1)+[R_1-(n^2+2n-2)]\sum_{r=1}^{m} s_r^{-1},\\
&\mu_3=\kappa_3=8m(n-1)+2[11R_1-(9n^2+18n-16)]\sum_{r=1}^{m} s_r^{-1}\\
&\qquad\qquad +[R_2-(3n+22)R_1+2(n^3+9n^2+14n-12)]\sum_{r=1}^{m} s_r^{-2},\\
&\kappa_4=48m(n-1)+96[4R_1-(3n^2+6n-5)]\sum_{r=1}^{m} s_r^{-1}\\
&\qquad\qquad +8[14R_2-2(14n+83)R_1+3(5n^3+41n^2+62n-48)]\sum_{r=1}^{m} s_r^{-2}\\
&\qquad\qquad +[R_3-4(n+28)R_2-3R_1^2+4(3n^2+56n+236)R_1\\
&\qquad\qquad\qquad -6(n^4+20n^3+116n^2+160n-120)]\sum_{r=1}^{m} s_r^{-3},\\
&\mu_4=\kappa_4+3\kappa_2^2.
\end{aligned}\right\} \qquad \ldots\ldots(5)$$

5. Case of $(n\times 2)$-fold classification (n degrees of freedom)

These are the most general formulae arrived at in this paper. Special cases analogous to equations (1) and (3) obviously arise. Only the latter will be given as it is used by Grüneberg and Haldane (1937). For n samples each consisting of s members, and each divided into two classes, whose expected values are ps and qs, we have

$$\chi^2=\sum_{r=1}^{n}\frac{(a_r-ps)^2}{spq}, \text{ degrees of freedom}=n,$$

where a_r is observed frequency in the first class of rth sample. Further, writing $k=(pq)^{-1}$ we have for the moments and cumulants of χ^2:

$$\mu_1'=\kappa_1=n,$$

$$\mu_2=\kappa_2=ns^{-1}(2s+k-6),$$

$$\mu_3=\kappa_3=ns^{-2}[8s^2+(22k-112)s+k^2-30k+120],$$

$$\mu_4=ns^{-3}[12(n+4)s^3+12\{(k-6)n+8(4k-19)\}s^2 + \{3n(k-6)^2+16(7k^2-125k+420)\}+k^3-126k^2+1680k-5040],$$

$$\kappa_4=ns^{-3}[48s^3+96(4k-19)s^2+16(7k^2-125k+420)s + k^3-126k^2+1680k-5040]. \qquad \text{.....(6)}$$

Hence $$\beta_1=\frac{[8s^2+(22k-112)s+k^2-30k+120]^2}{ns(2s+k-6)^3}.$$

It will be seen that when k is large compared with s, that is to say when one of the expectations is a small fraction, β_1 approximates to k/ns, whereas its value when s tends to infinity is $8/n$. But for moderate values of s the skewness may be considerably less than in the classical case.

Two numerical values of k are important in genetics. If $p=q=\frac{1}{2}$, $k=4$, and:

$$\left.\begin{aligned}\mu_2=\kappa_2&=2n(s-1)s^{-1},\\ \mu_3=\kappa_3&=8n(s-1)(s-2)s^{-2},\\ \kappa_4&=16n(s-1)(3s^2-15s+17)s^{-3},\\ \beta_1&=\frac{8(s-2)^2}{ns(s-1)}.\end{aligned}\right\} \qquad \text{.....(7)}$$

Thus for example when $n=50$, if $s=4$, $\beta_1=\cdot053$, whereas when s is infinite $\beta_1=\cdot16$.

When $p=\frac{1}{4}$, $q=\frac{3}{4}$, $k=\frac{16}{3}$, we have:

$$\left.\begin{aligned}\mu_2=\kappa_2&=\frac{2n(3s-1)}{3s},\\ \mu_3=\kappa_3&=\frac{8n(9s^2+6s-13)}{9s^2},\\ \kappa_4&=\frac{16n(81s^3+378s^2-1284s+823)}{27s^2},\\ \beta_1&=\frac{8(9s^2+6s-13)^2}{3ns(3s-1)^3}.\end{aligned}\right\} \qquad \text{.....(8)}$$

For example if $n=50$, $s=4$, $\beta_1=\cdot24$.

The values of s will generally vary from one sample to another. In this case the values of the cumulants are the sums of the values found for the different sample sizes.

6. CASE OF n-FOLD CLASSIFICATION (n DEGREES OF FREEDOM)

A limiting case of the $(n \times 2)$-fold classification arises when the values of s_r tend to infinity, but the expectations, $g_r = ps_r$, in say the first category remain finite. If a_r is the observed frequency in this category, it may be regarded as resulting from a single sample from a Poisson series with expected value g_r. Then

$$\chi^2 = \sum_{r=1}^{n} \frac{(a_r - g_r)^2}{g_r},$$

and using equations (4) we have

$$\begin{aligned}
&\overline{\chi^2} = n, \\
&\mu_2 = \kappa_2 = 2n + \Sigma g_r^{-1}, \\
&\mu_3 = \kappa_3 = 8n + 22\Sigma g_r^{-1} + \Sigma g_r^{-2}, \\
&\kappa_4 = 48n + 384\Sigma g_r^{-1} + 112\Sigma g_r^{-2} + \Sigma g_r^{-3}.
\end{aligned}$$

If we call s the size of the whole sample, and let $g_r = sp_r$, while $R_1 = \Sigma p_r^{-}$, $R_2 = \Sigma p_r^{-2}$, $R_3 = \Sigma p_r^{-3}$, we can write, in full analogy with equations (2):

$$\left.\begin{aligned}
&\mu_1' = \kappa_1 = n, \\
&\mu_2 = \kappa_2 = 2n + s^{-1}R_1, \\
&\mu_3 = \kappa_3 = 8n + 22s^{-1}R_1 + s^{-2}R_2, \\
&\mu_4 = 12n(n+4) + 12(n+32)s^{-1}R_1 + s^{-2}(3R_1^2 + 112R_2) + s^{-3}R_3, \\
&\kappa_4 = 48n + 384s^{-1}R_1 + 112s^{-2}R_2 + s^{-3}R_3.
\end{aligned}\right\} \quad \ldots(9)$$

The great simplicity of these expressions as compared with (2) is noteworthy. The extra terms in (2) represent diminutions in the moments due to the loss of one degree of freedom.

7. A WEIGHTING CORRECTION

If we have a number of samples of different sizes s_r their variances will differ. Now when s is large the probability that each sample will make a given contribution to χ^2 is equal. If we wish to reinstate this condition as far as possible, we must arrange for a proper weighting of the contributions made from the various samples.

If we have m samples, in each of which there are n classes in expected numbers sp_1, sp_2, etc., s being the number in the sample, taking the mean and variance from equation (2), we put for each sample:

$$\zeta = \frac{s^{\frac{1}{2}}[\chi^2 - (n-1)]}{[2(n-1)s - (n^2 + 2n - 2) + \Sigma p_i^{-1}]^{\frac{1}{2}}}.$$

Then in each sample the mean of ζ is 0, and its variance 1. Hence for the m

samples the variance of the sum of the ζ's is 0 and its variance n. In the case of a $(n \times 2)$-fold table with n degrees of freedom

$$\zeta = \frac{\chi^2 - 1}{\left(2 + \frac{k-6}{s}\right)^{\frac{1}{2}}}.$$

This weighting correction is hardly worth making in the Mendelian cases where $p = \frac{1}{2}$, $k = 4$, and $p = \frac{1}{4}$, $k = \frac{16}{3}$. Here the weighting factors vary from ·7071 for $s = \infty$ to 1 for $s = 2$ in the case when $k = 4$, and from ·7071 when $s = \infty$ to ·866 when $s = 1$ in the case when $k = \frac{16}{3}$. But when k is larger the variation is considerable. Thus, in the case of a selfed autotetraploid $p = \frac{1}{36}$, and the weighting factor falls from ·7071 when s is infinite to ·1740 when $s = 1$.

8. Discussion

The results obtained for the moments of χ^2 in a $(n \times 2)$-fold table when p is fixed do not agree with those given by Cochran (1936), who finds a mean value $n-1$, and a variance $2(n-1) + \frac{(n-1)^2(k-6)}{ns}$. These results only differ from those here given by a factor of the order $1 - \frac{1}{n}$, and are therefore satisfactory when n is large. However, when p is known and does not have to be estimated from the data there are clearly n degrees of freedom, and not $n-1$; hence my own results would appear to be slightly more accurate than Cochran's. I have, however, no reason to doubt the accuracy of Cochran's results when p is estimated from the totals. It is noteworthy that while in the cases considered here, where χ^2 is used as a test of goodness of fit, its mean is always exactly equal to the number of degrees of freedom, this is no longer so when it is used as a test of homogeneity. Thus Cochran finds for a $(n \times 2)$-fold table with $n-1$ degrees of freedom, a mean χ^2 of $n - 1 + \frac{n-1}{ns} + \frac{1}{ns^2}$, when all samples contain s members.

It follows from the results here given that the distribution of χ^2 for large values of n generally approximates fairly closely to normality. It would of course be possible to find a function of χ^2 whose distribution is much more nearly normal. Thus Wilson and Hilferty (1931) found that, when s is large,

$$\left[\left(\frac{\chi^2}{n}\right)^{\frac{1}{3}} - 1 + \frac{2}{9n}\right]\left(\frac{9n}{2}\right)^{\frac{1}{2}}$$

is very nearly normally distributed with mean zero and variance unity. It may also be desirable, as Fisher (1922), Neyman and Pearson (1928) and Cochran (1936) point out, to use the logarithm of the likelihood of the sample, rather than χ^2, as a test of goodness of fit when expectations are small. It is, however, worth

pointing out that, in the estimation of the frequency of lethal genes in autosomes, a problem with which I hope to deal later, χ^2 appears to furnish a simple and satisfactory estimate, and its distribution must therefore be known.

9. SUMMARY

Exact expressions have been found for the mean and the first four moments of χ^2 in cases where it is used as a test of goodness of fit, that is to say in $(m \times n)$-fold tables with $m(n-1)$ degrees of freedom. The mean is always exactly $m(n-1)$. The expressions for the higher moments are more complicated. Information has therefore been obtained which will make it possible to apply the χ^2 test without restriction on the size of the samples on the numbers expected. The results do not apply where χ^2 is used as a test of homogeneity, the expectations being deduced from observed totals.

REFERENCES

COCHRAN, W. G. (1936). "The χ^2 distribution for the binomial and Poisson series, with small expectations." *Ann. Eugen.*, Camb., VII, 207.

FISHER, R. A. (1922). "On the mathematical foundations of theoretical statistics." *Philos. Trans.* A, CCXXII, 309–68.

GRÜNEBERG, H. and HALDANE, J. B. S. (1937). "Test of goodness of fit applied to records of Mendelian segregation in mice." *Biometrika*, XXIX, 144–53.

NEYMAN, J. and PEARSON, E. S. (1928). "On the use and interpretation of certain test criteria for purposes of statistical inference. Part II." *Biometrika*, XX A, 263–94.

PEARSON, K. (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." *Phil. Mag.* L, 157–75.

PEARSON, K. (1932). "Experimental Discussion of the (χ^2, P) test for goodness of fit." *Biometrika*, XXIV, 351–81.

WILSON and HILFERTY (1931). *Nat. Acad. Sci.* XVII, No 12, 684.

TESTS OF GOODNESS OF FIT APPLIED TO RECORDS OF MENDELIAN SEGREGATION IN MICE

By HANS GRÜNEBERG and J. B. S. HALDANE

Department of Genetics, University College, London

It has long been known that in some, though not in all instances, the totals in cases of segregation involving a small number of genes were in satisfactory accord with the simple numerical ratios expected according to Mendel's laws. Divergences could often be explained by selective mortality of zygotes between the time of fertilization and the time when the characters in question could be determined.

But where the totals were in agreement with Mendelian expectation it was not clear that individual families might not show an unexpected number, either larger or smaller than that expected on sampling theory, of large deviations from the expected ratio.

Where the families, and the expectations in all groups, were sufficiently large, it was possible to apply Pearson's classical χ^2 method (e.g. de Winton & Haldane, 1933). Where samples were smaller this was no longer possible. However, Haldane (1937, pp. 133–43 above) has calculated the moments of the distribution of χ^2 when expectations are small.

If we are dealing with n samples from a population in which two classes occur with frequency p and $1-p$, if s_r be the size of the rth sample, then the principal parameters of the distribution of χ^2 are given in Table I for the two Mendelian cases where $p = \frac{1}{2}$ and $\frac{1}{4}$. These results follow if we put $m = n$, $n = 2$ in Haldane's equation (5) (p. 139 above).

TABLE I

Distribution parameters of χ^2 in Mendelian cases

	$p=\frac{1}{2}$	$p=\frac{1}{4}$
Mean	n	n
Variance	$2(n-\Sigma s_r^{-1})$	$\frac{3}{2}(3n-\Sigma s_r^{-1})$
$\beta_1=\gamma_1^2$	$\frac{8(n-3\Sigma s_r^{-1}+2\Sigma s_r^{-2})^2}{(n-\Sigma s_r^{-1})^3}$	$\frac{8(9n+6\Sigma s_r^{-1}-13\Sigma s_r^{-2})^2}{3(3n-\Sigma s_r^{-1})^3}$
$\gamma_2=\beta_2-3$	$\frac{4(3n-18\Sigma s_r^{-1}+32\Sigma s_r^{-2}-17\Sigma s_r^{-3})}{(n-\Sigma s_r^{-1})^2}$	$\frac{4(81n+378\Sigma s_r^{-1}-1284\Sigma s_r^{-2}+823\Sigma s_r^{-3})}{3(3n-\Sigma s_r^{-1})^2}$

In the case of our data it will be shown that n is often so large that β_1 (or γ_1) and γ_2 are of the order of their own sampling errors or less. The distribution of χ^2 can thus be treated as approximately normal.

Our experimental data are as follows:

(1) Records of 562 litters including 3707 mice, from back-crosses of mice heterozygous for the normal colour gene **C** and its allelomorph $\mathbf{c^d}$ to recessives. These were obtained during an experiment (Grüneberg, 1936) on the linkage of **C** with two other genes. The ratios obtained for these two genes are closely correlated with those for **C** and $\mathbf{c^d}$, and hence will not be given, as they do not give independent information.

(2) Records of 273 litters including 1366 mice, from a back-cross of $\mathbf{Cc^d} \times \mathbf{c^dc^d}$ and reciprocally. These are selected. In many cases the full-coloured parent was not known to be heterozygous for $\mathbf{c^d}$. Families of less than 7 were rejected if the full-coloured parent was not known to be heterozygous. (By a family is meant a group of litters from the same two parents.)

(3) Records of 243 litters including 1198 mice, from matings $\mathbf{Cc^d} \times \mathbf{Cc^d}$. Here again families of less than 16 individuals were rejected unless both parents were known to be heterozygous.

(4) Records of 226 litters including 1279 mice, obtained by Fisher & Mather (1936) in the course of a linkage experiment, and very kindly put at our disposal by the authors. These litters were derived from matings of mice heterozygous for five or six genes, with multiple recessives. One of these genes (for recessive light head) was not recorded in all litters, and gave aberrant ratios. The other five, and sex, were recorded in all litters (except blue dilution in one). The records given to us cover 50 mice beyond those on which Fisher & Mather's published results are based.

All these data, totalling 7550 mice in 1304 litters, are collected in Table II. The first 562 litters are divided into five groups (I, II, III, IV and V) representing different experiments (see Grüneberg, 1936). The total is also given. The table is to be read as follows. The first column gives litter size. The second two, headed **D** and **r**, give distribution as between dominants and recessives, or in the case of (4) for sex, as between males and females. Subsequent columns give the numbers of litters of this type in the various experiments.

Table III gives the totals in the various experiments, each χ^2 having 1 degree of freedom. In only two cases (1, III and 4, blue) does the deviation exceed twice the standard error. If the data of Exp. 1 are combined and a single χ^2 found for them, the total χ^2 for the four experiments is 10·87 for 9 degrees of freedom, a very moderate value. If they are considered separately, $\chi^2 = 21{\cdot}86$ for 13 degrees of freedom. P now just exceeds 0·05, which is generally taken as the criterion of significance.

TABLE II

Litter size	D	r	1						2	3	4					
			I	II	III	IV	V	Total			Sex	a	wv	s	b	d
1	1	0	1	2	—	1	1	5	4	11	2	—	2	2	2	2
	0	1	—	4	—	2	—	6	7	3	2	4	2	2	2	2
2	2	0	—	—	—	—	1	1	7	10	1	3	3	2	3	1
	1	1	5	2	4	—	3	14	16	9	4	6	5	5	6	7
	0	2	2	4	—	—	—	6	5	2	6	2	3	4	2	3
3	3	0	3	2	—	—	—	5	3	17	1	4	4	3	4	7
	2	1	3	1	5	1	—	10	13	14	10	7	11	11	5	6
	1	2	5	3	4	4	3	19	17	1	12	8	5	10	9	7
	0	3	3	—	1	1	1	6	5	1	1	5	4	—	6	4
4	4	0	1	—	3	—	—	4	—	13	1	2	1	—	2	2
	3	1	2	1	3	—	1	7	11	14	11	7	5	8	9	8
	2	2	9	2	3	1	1	16	11	7	9	11	11	13	11	10
	1	3	5	4	1	—	1	11	12	1	8	6	7	6	4	7
	0	4	1	—	1	—	2	4	1	1	—	3	5	2	3	2
5	5	0	1	—	—	—	—	1	3	9	1	—	—	1	—	—
	4	1	4	2	5	1	—	12	2	12	8	5	5	5	6	6
	3	2	8	1	9	2	2	22	21	7	15	15	18	14	14	15
	2	3	6	5	3	1	3	18	9	8	11	17	10	12	16	11
	1	4	2	3	3	—	—	8	8	—	5	6	8	9	6	10
	0	5	3	—	2	—	—	5	—	—	3	—	2	2	1	1
6	6	0	—	—	—	—	—	—	1	6	—	—	—	1	1	—
	5	1	4	—	6	2	—	12	5	18	5	2	5	3	1	5
	4	2	3	3	6	2	1	15	8	7	10	5	8	8	6	10
	3	3	9	1	9	3	2	24	18	7	8	11	14	9	13	12
	2	4	7	3	6	3	2	21	6	—	7	13	4	9	7	5
	1	5	2	1	2	3	3	11	3	—	3	2	2	1	3	1
	0	6	—	2	—	—	—	2	2	—	—	—	—	2	2	—
7	7	0	—	—	—	—	—	—	—	8	—	—	—	—	2	1
	6	1	1	1	3	1	2	8	2	10	2	1	2	2	—	2
	5	2	5	1	1	—	1	8	9	9	5	5	5	8	4	7
	4	3	6	—	8	2	1	17	14	9	8	7	9	6	9	8
	3	4	5	2	7	3	5	22	8	3	11	8	7	7	6	7
	2	5	2	2	—	4	1	9	8	1	2	8	6	5	6	4
	1	6	1	—	—	—	—	1	3	—	1	1	1	2	3	1
	0	7	—	—	—	—	—	—	—	—	1	—	—	—	—	—

TABLE II (*continued*)

Litter size	D	r	1						2	3	4					
			I	II	III	IV	V	Total			Sex	a	wv	s	b	d
8	8	0	—	—	—	1	—	1	—	3	1	—	—	—	—	—
	7	1	1	—	2	—	1	4	—	3	1	1	2	1	1	1
	6	2	2	—	2	—	4	8	2	7	5	3	5	3	8	1
	5	3	4	5	10	4	4	27	3	4	8	4	6	9	5	12
	4	4	6	—	9	1	3	19	5	—	7	9	12	10	5	9
	3	5	6	2	6	2	6	22	4	—	5	12	4	6	9	6
	2	6	—	1	3	2	2	8	3	—	5	3	2	3	3	3
	1	7	2	—	1	3	—	6	—	—	—	—	1	—	1	—
	0	8	—	—	—	—	1	1	—	—	—	—	—	—	—	—
9	9	0	—	—	1	—	—	1	1	—	—	—	—	—	—	—
	8	1	—	—	—	1	—	1	—	—	—	—	—	—	—	—
	7	2	2	1	2	1	2	8	—	1	—	1	—	—	2	1
	6	3	5	1	3	5	2	16	3	3	4	2	6	2	7	5
	5	4	3	1	3	4	2	13	4	—	6	4	6	2	2	5
	4	5	2	1	4	2	3	12	2	—	4	4	2	7	3	3
	3	6	2	—	4	5	4	15	1	1	1	3	1	3	—	1
	2	7	4	1	1	2	1	9	1	—	—	1	—	1	1	—
	1	8	1	—	—	—	—	1	1	—	—	—	—	—	—	—
	0	9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
10	9	1	—	—	—	—	—	—	—	1	—	—	—	—	—	—
	8	2	—	—	—	3	1	4	—	1	—	—	—	—	1	—
	7	3	6	—	2	2	—	10	—	—	1	—	—	—	—	1
	6	4	5	—	2	1	3	11	—	—	—	—	1	1	3	1
	5	5	5	1	—	1	—	7	—	1	1	1	1	1	—	1
	4	6	1	1	—	2	1	5	—	—	1	1	2	1	—	—
	3	7	2	1	2	2	—	7	—	—	1	1	—	1	—	—
	2	8	1	—	—	—	1	2	—	—	—	1	—	—	—	—
	1	9	—	—	—	1	1	2	—	—	—	—	—	—	—	—
11	9	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	8	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	7	4	—	—	—	—	—	—	—	—	—	—	—	1	—	—
	6	5	1	—	—	—	—	1	1	—	—	1	—	—	—	1
	5	6	2	—	—	—	—	2	—	—	—	—	1	—	—	—
	4	7	1	—	—	1	—	2	—	—	1	—	—	—	1	—
	3	8	1	—	—	—	—	1	—	—	—	—	—	—	—	—
	2	9	1	—	—	—	—	1	—	—	—	—	—	—	—	—
12	9	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	8	4	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	7	5	—	—	—	1	—	1	—	—	—	—	—	—	—	—
	6	6	—	—	—	1	1	2	—	—	—	—	—	—	—	—
	5	7	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	4	8	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	3	9	—	—	—	—	1	1	—	—	—	—	—	—	—	—
14	6	8	—	—	—	—	1	1	—	—	—	—	—	—	—	—

TABLE III

Exp.	Dominants	Recessives	Expectation	χ^2
1, I	567	577	572	0·087
1, II	162	198	180	3·600
1, III	528	457	492·5	5·118
1, IV	306	326	316	0·645
1, V	277	309	293	1·747
1, Total	1840	1867	1853·5	0·197*
2	685	681	683	0·012
3	899	299	299·5	0·001
4, sex	662	617	639·5	1·583
4, agouti	608	671	639·5	3·103
4, brown	648	631	639·5	0·226
4, spotting	636	643	639·5	0·038
4, wavy	658	621	639·5	1·070
4, blue	673	596	634·5	4·636

* This value of χ^2 is not the total of the five values given above it, but the value, having 1 degree of freedom, calculated from the totalled dominants and recessives of Exp. 1.

In all the cases the expectations are, of course, so large that we can use the classical χ^2 with complete confidence.

The calculation of χ^2 for each experiment from the data of Table II is rapid and simple. If a and b are the numbers of dominants and recessives, then in the case of a back-cross, where equality is expected, we multiply the numbers of litters containing a dominants and b recessives by $(a-b)^2$, sum the products for each value of s, and divide the sum by s. This gives the contribution to χ^2 made by litters of that particular size.

For example in the case of the total of Exp. 1 and litters of 6 mice the calculations are as follows:

Dominants	Recessives	$(a-b)^2$	n	$s\chi^2$
6	0	36	0	0
5	1	16	12	192
4	2	4	15	60
3	3	0	24	0
2	4	4	21	84
1	5	16	11	176
0	6	36	2	72
			85	584

Hence for litters of $s = 6$, $\chi^2 = \frac{584}{6} = 97{\cdot}\dot{3}$, the expected value as a result of random sampling being 85.

The results of applying this method to the data of Exp. 1 are given in Table IV, and compared with expectations in Table V. It will be seen that in every case χ^2

TABLE IV

	1, I		1, II		1, III		1, IV		1, V		1, Total	
s	n	χ^2	n	χ^2	n	χ^2	n	χ^2	n	χ^2	n	χ^2
1	1	1·000	6	6·000	0	—	3	3·000	1	1·000	11	11·000
2	7	4·000	6	8·000	4	0·000	0	—	4	2·000	21	14·000
3	14	20·667	6	7·333	10	6·000	6	4·667	4	4·000	40	42·667
4	18	15·000	7	5·000	11	20·000	1	0·000	5	10·000	42	50·000
5	24	33·600	11	10·200	22	26·800	4	2·400	5	1·000	66	74·000
6	25	22·667	10	18·667	29	29·333	13	16·667	8	10·000	85	97·333
7	20	17·714	6	7·714	19	14·143	10	9·429	10	10·571	65	59·571
8	21	22·500	8	5·500	33	31·500	13	28·500	21	29·500	96	117·500
9	19	29·667	5	6·778	18	25·111	20	24·444	14	14·889	76	100·889
10	20	18·800	3	2·000	6	7·200	12	24·800	7	15·200	48	68·000
11	6	7·818	0	—	0	—	1	0·818	0	—	7	8·636
12	0	—	0	—	0	—	2	0·333	2	3·000	4	3·333
13	0	—	0	—	0	—	0	—	0	—	0	—
14	0	—	0	—	0	—	0	—	1	0·286	1	0·286
	175	193·433	68	77·192	152	160·087	85	115·058	82	101·446	562	647·215

TABLE V

Exp.	χ^2	n	χ^2-n	σ	d/σ
1, I	193·43	175	+18·43	16·87	+1·09
1, II	77·19	68	+ 9·19	9·96	+0·92
1, III	160·09	152	+ 8·09	15·83	+0·51
1, IV	115·06	85	+30·06	10·96	+2·74
1, V	101·45	82	+19·45	11·62	+1·67
1, Total	647·22	562	+85·22	30·12	+2·83

exceeds its expectation, that the excess is significant in the total and in 1, IV, and is very probably so in 1, V.

The effect of the corrections to Pearson's χ^2 is of interest. The variance of the total is reduced from $2n$, or 1124, to $2(n-\Sigma s_r^{-1})$ or 907·46. Thus σ is reduced from 33·52 to 30·12, and $\frac{\chi^2-n}{\sigma}$ is increased from 2·54 to 2·83. The normality of the distribution of χ^2 is also improved. β_1 is reduced from 0·0258 to 0·0079, and γ_2 from 0·0387 to 0·0097. The closeness of the approach to normality may be realized as follows. The variance of $\gamma_1 = \beta_1^{\frac{1}{2}}$ in a sample of N from a normal population is approximately $6/N$. Hence a value of $\beta_1 = 0{\cdot}0079$ would be found about once in three times in a sample of $6/\beta_1$ or 760 individuals from a normal population. It is in fact negligible.

In Exp. 1, II, which gave only 68 litters, $\beta_1 = 0{\cdot}0602$, $\beta_2 = 3{\cdot}00145$. The skewness is hardly worth considering in a test of significance.

Exps. 2 and 3 give the results shown in Table VI. To calculate the values of χ^2 in Exp. 3 we note that for a litter of s containing a dominants and b recessives

$$\chi^2 = \frac{(\frac{3}{4}s-a)^2}{\frac{3}{4}s} + \frac{(\frac{1}{4}s-b)^2}{\frac{1}{4}s} = \frac{(a-3b)^2}{3s}.$$

TABLE VI

	Exp. 2		Exp. 3	
s	n	χ^2	n	χ^2
1	11	11·000	14	12·667
2	28	24·000	21	24·667
3	38	34·000	33	30·333
4	35	27·000	36	44·000
5	43	39·000	36	46·133
6	43	48·667	38	31·556
7	44	42·857	40	53·714
8	17	13·500	17	12·667
9	13	21·889	5	9·370
10	0	0·000	3	4·667
11	1	0·091	0	0·000
Total	273	262·004	243	269·774

Hence if we multiply the number of each litter type by $(a-3b)^2$ and divide the total for each litter size by $3s$ we obtain the contribution of that litter size to χ^2.

The deviation in Exp. 2 is $-10{\cdot}996$, its standard error being 20·04. The deviation in Exp. 3 is $+26{\cdot}77$, its standard error being 18·89. Thus neither deviation is significant. We shall later have to consider the effect on χ^2 of selecting our material.

The results of applying the χ^2 test to Exp. 4 are given in Table VII. It will be seen that five out of the six values of χ^2 are less than their expectation. The variance is 353·41 except in the case of blue dilution, where it is 351·61. It will be seen that none of the deviations, taken by itself, is significant. The total value of χ^2 is 1277·38, its expectation being $1355 \pm 46{\cdot}03$. The deviation is $-1{\cdot}69$ times the standard error, which again is not significant. A considerably larger negative deviation would have suggested that the authors had suppressed a few aberrant families. An application of χ^2 to certain published work would, we are inclined to believe, give ground for such a suggestion.

We must next ask whether the large positive deviations of Exp. 1 can be explained. In order to analyse this experiment further the mice were grouped,

TABLE VII

s	n	χ^2					
		Sex	Non-agouti	Wavy	Spotting	Brown	Blue
1	4	4·000	4·000	4·000	4·000	4·000	4·000
2	11	14·000	10·000	12·000	12·000	10·000	8·000
3	24	13·333	32·000	10·667	16·000	34·667	37·333
4	29	23·000	33·000	36·000	22·000	33·000	31·000
5	43	48·600	26·200	39·000	45·400	32·600	39·000
6	33	32·667	22·667	26·667	40·000	37·333	26·000
7	30	29·429	26·000	27·143	32·857	39·714	34·000
8	32	39·000	24·500	32·500	24·000	38·000	20·000
9	15	6·111	11·444	7·889	8·778	15·889	9·667
10	4*	3·600	5·600	1·200	2·400	4·800	2·000
11	1	0·818	0·091	0·091	0·818	0·818	0·091
Total	226	214·558	195·502	197·157	208·253	250·821	211·091
d		−11·44	−30·50	−28·84	−17·75	+24·82	−13·91
d/σ		− 0·61	− 1·62	− 1·53	− 0·94	+ 1·32	− 0·74

* One family of 10 was not scored for blue dilution.

not in litters, but in families. The result is shown in Table VIII. The numbers of families, except in Exp. 1, I, and the total, are so small that the distribution of χ^2 is far from normal. And some of them are so small that the classical distribution is also inapplicable. It is clear, however, that the total χ^2 exceeds

TABLE VIII

Exp. 1, *litters grouped in families*

Exp.	Number of families	Number of mice	χ^2	χ^2-n	σ
1, I	107	1144	126·58	+19·58	13·56
1, II	26	360	33·03	+ 7·03	6·70
1, III	64	985	90·84	+26·84	10·72
1, IV	31	632	46·22	+15·22	7·54
1, V	25	586	33·74	+ 8·74	6·79
1, Total	253	3707	330·40	+77·40	21·13

its expectation by 3·66 times its standard error, and the divergences in Exps. 1, III and 1, IV must probably be regarded as significant. In each case, however, the divergence was mainly due to a single family. A family of 9 dominants and no recessive contributed 9 to the χ^2 of 1, III, and a family of 3 dominants and

14 recessives contributed 7·12 to the χ^2 of 1, IV. Were these families omitted no single experiment would yield a significant result, though their total would do so.

It is notable that when the litters are grouped by families, although n is reduced to 45 % of its former value, the excess of χ^2 is only reduced from 85·2 to 77·4. Thus there is little heterogeneity due to divergence of litters within families. We must look for a cause which affects families rather than litters. The following are possible causes:

(1) The presence of recessive lethal or sublethal genes linked with those segregating.

(2) The presence of monozygotic twins or multiplets.

(3) The existence of environmental factors which on some occasions favoured coloured mice, on others dilute mice, although on the whole they favoured neither.

(4) Abnormalities of meiosis leading to production of gametes in unequal numbers where equality was expected.

Only the first, and possibly the fourth of these, would affect families rather than litters. This is plausible on general grounds. The presence of the same lethal or sublethal gene in both parents is only likely as the result of inbreeding. There was inbreeding in all five parts of Exp. 1, but a perusal of Grüneberg's (1936) paper makes it clear that it was least in 1, I and 1, II, greater in 1, III, and greatest in 1, IV, and 1, V. Our results do not prove the segregation of lethal genes, but they are consistent with it. On the other hand such genes must have been rare or absent in Exps. 2, 3 and 4, although 2 and 3 involved a good deal of inbreeding and 4 a certain amount. It is hoped, by the application of this method to man, to obtain at least an upper limit to the frequency of lethals in human chromosomes, and possibly to obtain evidence suggesting their presence.

We must next consider the effect of selection on Exps. 2 and 3. In Exp. 2 the fully coloured parents were not always known to be heterozygous, and families of less than 7 were excluded unless the parent was known to be so. We may, however, have neglected some families of 7 or over derived from a heterozygous parent because they contained no dilute animals. The probability of such a family is 2^{-7} or less. Thirty-six families of 7 or more were inferred to be derived from a heterozygous parent because they contained one or more extreme dilute mice. The probability that such a group taken at random from the progeny of a heterozygote and a homozygote would have contained a family with no extreme dilute mice is 0·0365. A family of 7 would have contributed 7 to χ^2. Thus the value of χ^2 should be increased by about 0·25 of a unit to compensate for the effect of selection.

Similarly in Exp. 3 there were 30 families ranging from 16 to 35 in number which were inferred to be derived from matings of two heterozygotes because they contained at least one extreme dilute mouse. The probability that a family of n derived from two heterozygotes will include no recessive is $(\frac{3}{4})^n$. This is equal

to 0·0100 or less. Among 30 families of the given sizes the probability that one has been excluded on this ground is 0·0838. Such a family would contribute fairly heavily to χ^2. For example a family of 18 would contribute 6. It is concluded that the selection practised has probably reduced the value of χ^2 by less than half a unit. The data have therefore been legitimately used.

SUMMARY

The χ^2 test has been applied to data on Mendelian segregation on 1304 mouse litters containing 7550 mice. As some litters were scored for as many as six characters we have effectively 2433 litters, and 13935 mice. In 7 experiments χ^2 exceeded its expectation, in 6 it fell below it. None of the negative deviations was significant, but one of the positive ones was so. This is tentatively ascribed to the effect of recessive lethal genes in upsetting Mendelian segregation in an inbred population.

REFERENCES

DE WINTON, D. & HALDANE, J. B. S. (1933). "The genetics of *Primula sinensis*. II." *J. Genet.* XXVII, 1–44.

FISHER, R. A. & MATHER, K. (1936). "A linkage test with mice." *Ann. Eugen.*, Camb., VII, 265–80.

GRÜNEBERG, H. (1936). "Further linkage data on the albino chromosome of the house mouse." *J. Genet.* XXXIII, 255–65.

HALDANE, J. B. S. (1937). "The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small." *Biometrika*, XXIX, 133–43.

MISCELLANEA

(i) An Application of the Method of Maximum Likelihood

By WALTER A. HENDRICKS

Bureau of Animal Industry, United States Department of Agriculture

In a recent paper Pearson (1936) presented some rather critical observations on the interpretation of results derived from applications of the method of maximum likelihood. The author of the present paper has no desire to attempt a theoretical justification of the method of maximum likelihood. He can, however, add another scrap of evidence to the mounting total which has done much to justify the method by the principle of induction.

Suppose that a set of 17 individuals, drawn from some universe, may be divided into four classes. Let the numbers of individuals in the four classes be 3, 7, 2, and 5, respectively. Assume that a hypothesis regarding the universe causes us to expect twice as many individuals in the second class as in the first, and twice as many in the fourth class as in the third. In other words, assume that the respective probabilities for the four classes are p_1, $2p_1$, p_2, and $2p_2$. Let it be required to obtain estimates of p_1 and p_2 from the observed distribution of 17 individuals.

The investigator equipped with a knowledge of the more elementary aspects of simple sampling would doubtless proceed by reasoning that the probability of the occurrence of an individual in either the first or second class is equal to $3p_1$. He would equate this to the observed proportion, 10/17, of individuals in the two classes and solve the resulting equation, thus obtaining 10/51 or ·19607843 for his estimate of p_1. He would then apply the same process to the data in the third and fourth classes, or he would make use of the relation,

$$3p_1+3p_2=1, \qquad \text{.....(1)}$$

to reach the conclusion that p_2 is equal to 7/51 or ·13725490.

It is of interest to note that this solution, which would appeal to the experienced investigator, is exactly that to which the method of maximum likelihood leads.

In applying the method of maximum likelihood to the above problem, we are required to determine p_1 and p_2 so as to give the maximum value to the quantity, L, defined by

$$L=3\log p_1+7\log 2p_1+2\log p_2+5\log 2p_2, \qquad \text{.....(2)}$$

subject to the condition imposed by equation (1). This is equivalent to determining p_1, p_2, and λ_1 in such a manner as to give the maximum value to the quantity, Y, defined by

$$Y=10\log p_1+7\log p_2+\lambda_1(3p_1+3p_2-1). \qquad \text{.....(3)}$$

The required values of p_1, p_2, and λ_1 are given by the solution of the equations:

$$\left.\begin{aligned} 10/p_1+3\lambda_1&=0,\\ 7/p_2+3\lambda_1&=0,\\ 3p_1+3p_2&=1,\end{aligned}\right\} \qquad \text{.....(4)}$$

from which the values of p_1 and p_2 are found to be 10/51 and 7/51, respectively, exactly as before.

It is also of some interest to obtain estimates of p_1 and p_2 in such a manner that the familiar criterion of goodness of fit, χ^2, as applied to the comparison of observed and theoretical frequencies in the four classes, shall have its minimum value.

The value of χ^2 is given by the relation

$$\chi^2=9/17p_1+49/34p_1+4/17p_2+25/34p_2-17. \qquad \text{.....(5)}$$

To render this quantity a minimum, subject to the condition imposed by equation (1), we may determine p_1, p_2, and λ_2 in such a manner as to minimize the quantity, z, defined by

$$z = 67/p_1 + 33/p_2 + \lambda_2(3p_1 + 3p_2 - 1). \qquad \text{.....(6)}$$

The required values of p_1, p_2, and λ_2 are given by the solution of the equations,

$$\left.\begin{aligned} -67/p_1^2 + 3\lambda_2 &= 0, \\ -33/p_2^2 + 3\lambda_2 &= 0, \\ 3p_1 + 3p_2 &= 1, \end{aligned}\right\} \qquad \text{.....(7)}$$

from which the values of p_1 and p_2 are found to be ·19586990 and ·13746343, respectively.

These estimates of p_1 and p_2 differ very little from those obtained by the two preceding methods. However, they are different. This is not surprising, since the application of the χ^2 test to data such as those under consideration involves certain well-known approximations. The fact that the differences are rather small is in agreement with results obtained by Fisher (1934) in a similar comparison of methods of estimation.

The most interesting feature of the results presented in this paper is the fact that the method of maximum likelihood, as applied to the present problem, led to results which are in exact agreement with those obtained as natural consequences of the established theory of simple sampling.

REFERENCES

FISHER, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed. Edinburgh and London: Oliver and Boyd.

PEARSON, KARL (1936). "Method of moments and method of maximum likelihood." *Biometrika*, XXVIII, 34–59.

(ii) Maximum Likelihood and Methods of Estimation

BY E. S. PEARSON

IN the preceding Note Mr Walter A. Hendricks has given an example of the difference between the result of applying the methods of maximum likelihood and of minimum χ^2. The illustration is interesting and suggestive although I do not know whether it can be described as making any contribution towards *justifying* the general application of the method of maximum likelihood. When two or more alternative methods of estimation are available, we may choose between them in many ways, e.g. by an appeal to intuition or by an appeal to practical expediency. On both these counts the method of maximum likelihood would, in the present instance, score points over the method of minimum χ^2. But when we turn to the alternative methods of fitting frequency curves, with which my father was concerned in the paper referred to, e.g. (i) by maximum likelihood; (ii) by minimum χ^2; (iii) by moments, the choice is far less easy to make. From the point of view of any commonly experienced sense of intuition, I fancy there can be no unique answer; from the point of view of practical expediency, in many cases the moment method clearly wins.

As I have mentioned elsewhere in this *Journal*,* it is I think the practical worker who will give the final casting vote between alternative theoretical methods, basing his decision on considerations of practical utility. In the growing complexity of mathematical statistics it is, however, often difficult for him to make his choice without the aid of simple guiding principles which appeal to his intuition. The concept of maximum likelihood involves such a principle. To assign to an unknown probability, p, that value which, if it were correct, would make the occurrence of the observed result more likely than any other value, is a procedure which has a simple intuitional appeal. But clearly there may be other guiding principles which are equally useful to follow.

* Pp. 53–64 above.

Consider Mr Hendrick's illustration. It is almost certain that neither of the 8 decimal-place estimates of p_1 which he gives, ·1960,7843 and ·1958,6990, is the true population value. What would be of more use to the practical man than either of these would be a rule for obtaining from the sample an upper and lower estimate for p_1, say $\overline{p}_1$ and $\underline{p}_1$, such that the statement

$$\underline{p}_1 \leqslant p_1 \leqslant \overline{p}_1 \qquad \text{.....(1)}$$

might be made with a certain measure of confidence, expressed by the risk of error, say α, involved in the statement. In so far as the method permits the adjustment of the values of α and of the breadth of the interval $(\underline{p}_1, \overline{p})$, it would have a greater appeal than any method which only provides a single-valued estimate of p_1. If now two alternative methods are available, both enabling us to estimate an interval $(\underline{p}_1, \overline{p})$, we may make a choice between them, basing that choice upon a comparison of (*a*) the risk of error involved in the statement (1), and (*b*) the breadth of the intervals.

In this twofold principle we have gone beyond that involved in the simple choice of p_1 by maximizing the likelihood. That method may form a part of the process of determining the interval $(\underline{p}_1, \overline{p}_1)$, but it is now the means to an end, not the end itself.

When we come to the more complex problem of curve fitting, it is far less clear what result will be most useful to the practical worker. Starting from a given functional equation

$$y = p(x \mid \theta_1, \theta_2, \ldots, \theta_c) \qquad \text{.....(2)}$$

involving c unknown parameters θ, it is clear that a possible principle of estimation is to assign to the parameters the values, which if they were the population values, would make the occurrence of the observed sample more likely than any other set of parametric values. But many have felt that there is something remote about the abstract conception involved. If the maximum likelihood procedure leads to estimates T_i of θ_i $(i = 1, 2, \ldots, c)$ which in random sampling have smaller standard errors than any other form of estimates T_i, a result has been reached with a more direct practical appeal. The achievement would be greater still if a method were available for determining upper and lower limits $\overline{T}_i$ and $\underline{T}_i$ so that the statement

$$\underline{T}_i \leqslant \theta_i \leqslant \overline{T}_i (i = 1, 2, \ldots, c) \qquad \text{.....(3)}$$

could be made jointly with regard to the c parameters, with a given risk of error. But in the present state of development of the theory of maximum likelihood, as applied to the fitting of frequency curves to samples of *finite* size, can it be said that such results have been achieved?

It must also be remembered that in so far as the statistician wishes to use his frequency curve for graduation purposes, the agreement between observation and fitted curve throughout the range of significant frequency may make a more direct and simpler appeal to him than any information regarding the values of the parameters, θ_i. The quantity

$$\chi^2 = \sum_s \left(\frac{n_s - m_s}{\sqrt{m_s}} \right)^2, \qquad \text{.....(4)}$$

summing up the relative discrepancy, may be more closely correlated with his conception of goodness of fit than any measure based on the likelihood, L, or on the reliability of the estimated parameters.

Finally, the question of practical utility must play a dominating part; at present the method of fitting by moments, making where necessary certain empirical adjustments, does provide a practical working tool. Until far more exploratory work has been carried out on the application of the method of maximum likelihood in fitting frequency curves, it is quite impossible to attempt any assessment of its final value.

It was some of these considerations, expressed perhaps in a different form, that I believe my father had in mind when challenging the claim that the method of maximum likelihood is the only efficient method of fitting frequency curves.

(iii) A Note on Unbiased Limits for the Correlation Coefficient

By F. N. DAVID

We shall assume that a sample of size n has been randomly drawn from a normal bivariate population such as

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\xi_1)^2}{\sigma_1^2}+\frac{2\rho(x-\xi_1)(y-\xi_2)}{\sigma_1\sigma_2}+\frac{(y-\xi_2)^2}{\sigma_2^2}\right\}},$$

and that we are interested in ρ, the coefficient of correlation between x and y in the population. It is possible that we may require answers to two questions:

(1) Given the sample of size n, are these observations consistent with the hypothesis that $\rho = \rho_0$, where ρ_0 is some specified value?

(2) Given the sample of size n, how may we calculate ρ_1 and ρ_2, so that, subject to the risk which we are willing to undertake, the interval ρ_1 to ρ_2 will cover the true population value as often as possible?

There are other questions which we may ask, and these will be dealt with fully in the introduction to "Tables of the Ordinates and Probability Integral of the distribution of r" which

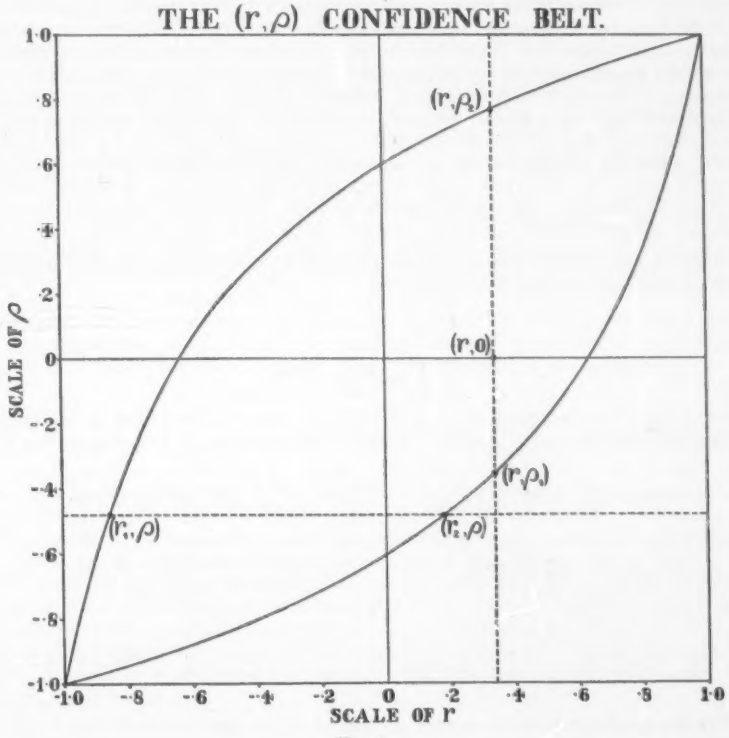


Fig. 1.

is shortly to be published. In the present note we shall confine ourselves solely to those questions in which a consideration of "bias" is important.

The methods of answering these questions may be illustrated with the help of the diagram on p. 157. If for a given n for each value of ρ in the range -1 to $+1$ we calculate from the probability distribution of r, say $p_n(r \mid \rho)$, limits r_1 and r_2 such that

$$\int_{-1}^{r_1} p_n(r \mid \rho)\, dr = \alpha_1, \quad \int_{r_2}^{+1} p_n(r \mid \rho)\, dr = \alpha_2, \qquad \text{.....(1)}$$

where
$$\alpha_1 = \alpha_2 \qquad \text{.....(2)}$$

and
$$\alpha_1 + \alpha_2 = \alpha = \text{constant}, \qquad \text{.....(3)}$$

then the points (r_1, ρ) and (r_2, ρ) will fall on two curves enclosing a lozenge-shaped belt as shown.* Accordingly when dealing with question (1), if we decide to reject the hypothesis tested, i.e. that $\rho = \rho_0$, whenever the point (r, ρ_0) falls outside the belt, we shall run a risk equal to α of rejecting the hypothesis when it is true. Question (2) may be answered with the aid of the same diagram. The point $(r, \rho = 0)$ is plotted and a line parallel to the axis of ρ is drawn through it. Suppose that this line cuts the belt in the points (r, ρ_1) and (r, ρ_2). Then we know that the interval ρ_1 to ρ_2 will cover the true value of ρ in the population in $100(1-\alpha)$ percentage of cases (Neyman, 1934). It will be noticed that the risk of error, α, would remain the same if equation (3) held, but not equation (2). Thus it is possible to obtain an infinite variety of belts satisfying (3) by following different principles in the determination of r_1 and r_2, or α_1 and α_2.

Neyman & Pearson (1936), when discussing questions similar to question (1), showed that in certain skew distributions the limits obtained by taking equal tail areas led to a curious anomaly. In such cases they found that the hypothesis tested was more likely to be rejected when it was true, than when in fact an alternative hypothesis was true. A test leading to such consequences Neyman & Pearson termed "biased". It is the object of the present note to find unbiased limits for r, and to compare them with the limits found by taking equal tail areas.

The probability distribution of r, for any n and ρ may be written as follows:

$$p_n(r \mid \rho) = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)!}\,\frac{(1-r^2)^{\frac{n-4}{2}}}{\pi}\,\frac{d^{n-2}}{d(r\rho)^{n-2}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right). \qquad \text{.....(4)}$$

Following the procedure of Neyman & Pearson we see that an unbiased test will be obtained by solving equations (5) and (6) for r_1 and r_2.

$$\int_{r_1}^{r_2} p_n(r \mid \rho)\, dr = 1-\alpha, \qquad \text{.....(5)}$$

$$\frac{d}{d\rho}\int_{r_1}^{r_2} p_n(r \mid \rho)\, dr = 0, \qquad \text{.....(6)}$$

where α is chosen, as is customary, according to the risk we are willing to undertake of rejecting the hypothesis as false when it is true. Differentiating (6) with respect to ρ we get

$$0 = \frac{-\rho(n-1)}{1-\rho^2}\int_{r_1}^{r_2}\frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)!}\,\frac{(1-r^2)^{\frac{n-4}{2}}}{\pi}\,\frac{d^{n-2}}{d(r\rho)^{n-2}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right)dr$$

$$+\int_{r_1}^{r_2}\frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)!}\,\frac{(1-r^2)^{\frac{n-4}{2}}}{\pi}\,r\cdot\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right)dr,$$

that is

$$0 = \frac{-\rho(n-1)(1-\alpha)}{(1-\rho^2)} + \int_{r_1}^{r_2}\frac{(1-\rho^2)^{\frac{n-1}{2}}}{(n-3)!}\,\frac{(1-r^2)^{\frac{n-4}{2}}}{\pi}\,r\cdot\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right)dr. \qquad \text{.....(7)}$$

* The diagram illustrates approximately the case $n = 10$, $\alpha = 0{\cdot}05$; charts of these "confidence belts" for varying n and α will be given in the publication referred to above.

Integrate (7) by parts, remembering that

$$\frac{d}{dr}\left[\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)\right]=\rho\frac{d^n}{d(r\rho)^n}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)$$

and we get

$$\int_{r_1}^{r_2}(1-r^2)^{\frac{n-4}{2}}r.\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)dr=-\left[\frac{(1-r^2)^{\frac{n-2}{2}}}{n-2}\right.$$

$$\left.\times\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)\right]_{r_1}^{r_2}+\frac{\rho}{n-2}\int_{r_1}^{r_2}(1-r^2)^{\frac{n-2}{2}}\frac{d^n}{d(r\rho)^n}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right)dr. \quad \text{.....(8)}$$

From (4) we see that we may write

$$p_{n+1}(r\mid\rho)=\frac{(1-\rho^2)^{\frac{n}{2}}}{(n-2)\,!}\frac{(1-r^2)^{\frac{n-3}{2}}}{\pi}\frac{d^{n-1}}{d(r\rho)^{n-1}}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right), \quad \text{.....(9)}$$

$$p_{n+2}(r\mid\rho)=\frac{(1-\rho^2)^{\frac{n+1}{2}}}{(n-1)\,!}\frac{(1-r^2)^{\frac{n-2}{2}}}{\pi}\frac{d^n}{d(r\rho)^n}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2r^2}}\right). \quad \text{.....(10)}$$

Substituting (10), (9) and (8) in (7) we get

$$(1-\alpha)=\int_{r_1}^{r_2}p_{n+2}(r\mid\rho)\,dr-\frac{\sqrt{1-\rho^2}}{\rho(n-1)}\left[p_{n+1}(r\mid\rho).\sqrt{1-r^2}\right]_{r_1}^{r_2}, \quad \text{.....(11)}$$

or making use of (5)

$$\int_{r_1}^{r_2}p_n(r\mid\rho)\,dr=\int_{r_1}^{r_2}p_{n+2}(r\mid\rho)\,dr-\frac{\sqrt{1-\rho^2}}{\rho(n-1)}\left[p_{n+1}(r\mid\rho).\sqrt{1-r^2}\right]_{r_1}^{r_2}. \quad \text{.....(12)}$$

Hence solving for r_1 and r_2 from (5) and (12) we should get the r_1 and r_2 which are the unbiased limits for r for a given n and ρ.

An algebraical solution of (5) and (12) proved elusive. Accordingly it was decided to solve equations (5) and (12) for r_1 and r_2 by means of trial and error, given one specific size of sample. The size of sample chosen was $n = 10$; this because it is unlikely that the correlation coefficient would be worked out for a sample of less than 10 observations. For all samples of more than 10 observations the bias, if any, would be expected to be less than that for the sample of 10, since the distribution curves of r tend slowly to normality.

The method of procedure was as follows: $1-\alpha$ was chosen as 0·95, and the first value of ρ to be considered was $\rho = 0{\cdot}5$. Using the unpublished tables (David, 1937) of the probability integral of r, r_1 and r_2 were found by backward interpolation into the tables, for

$$\alpha_1 = 0{\cdot}025 = \alpha_2,\ n = 10,\ \rho = 0{\cdot}5,\ \alpha = 0{\cdot}05.$$

Equal tail areas give $r_1 = -0{\cdot}1556$; $r_2 = 0{\cdot}8673$. The right-hand side of equation (11) was evaluated using these values of r_1 and r_2. Instead of being equal to 0·95, as it would have been had there been no bias, it was equal to 0·9512. Several other values of α_1 and α_2 were tried. Finally taking $\alpha_1 = 0{\cdot}0245$ and $\alpha_2 = 0{\cdot}0255$, by backward interpolation equation (5) gave $r_1 = -0{\cdot}1591$; $r_2 = 0{\cdot}8664$. Evaluating the right-hand side of (11), using these values for r_1 and r_2 it was found to equal 0·95. Hence the r_1 and r_2, given by $\alpha_1 = 0{\cdot}0245$ and $\alpha_2 = 0{\cdot}0255$, are the unbiased limits for r, and it is seen that the bias is very small.

The distribution curve of r for $\rho = 0$ is symmetrical, so in this case the limits r_1 and r_2 obtained by taking equal tail areas will also be the unbiased limits. The distribution curves of r gradually become asymmetrical as ρ increases. It therefore seems reasonable to suppose that the bias gradually increases with the asymmetry. We have investigated the case where $\rho = 0{\cdot}5$. Let us now consider the bias when $\rho = 0{\cdot}8$. The same method was carried out as before and we obtained

$$n = 10,\quad \rho = 0{\cdot}8\ \begin{cases}\alpha_1 = 0{\cdot}025 & \alpha_2 = 0{\cdot}025 & r_1 = +0{\cdot}4003 & r_2 = 0{\cdot}9550^+\\ \alpha_1 = 0{\cdot}0242 & \alpha_2 = 0{\cdot}0258 & r_1 = +0{\cdot}3959 & r_2 = 0{\cdot}9545^+.\end{cases}$$

We see that there is a greater difference between the tail areas than for $\rho = 0{\cdot}5$, but that for the r's the difference between the unbiased values and those obtained from equal tail area limits is only slightly increased for r_1 and slightly decreased for r_2. The area under the distribution curves for both $\rho = 0{\cdot}5$ and $\rho = 0{\cdot}8$ is unity, but the standard deviation for the curve $\rho = 0{\cdot}8$ is less than that for the curve $\rho = 0{\cdot}5$. Hence an alteration in our tail areas for the curve $\rho = 0{\cdot}8$, will mean much less change in the limits for r than for the same alteration in the tail areas of the curve for $\rho = 0{\cdot}5$. Our result accordingly seems reasonable, and we may therefore conclude that the unbiased limits for r follow very closely those limits which are found by taking equal tail areas.

In answering question (1) we see that in testing the hypothesis $\rho = \rho_0$, with admissible alternative hypotheses $-1 < \rho < \rho_0$ and $\rho_0 < \rho < +1$, if we used the limits for r obtained from equal tail areas we should reject the hypothesis tested as false according to the prescribed risk, but that there is a possibility that we should reject some other wrong hypothesis even less frequently.

In answering question (2) we may note that, between the points $(r_1, \rho = 0)$ and $(r_2, \rho = 0)$, the equal tail area interval for ρ is actually narrower than the unbiased interval, while for $(-1, \rho = 0)$ to $(r_1, \rho = 0)$, and $(r_2, \rho = 0)$ to $(+1, \rho = 0)$ the intervals practically coincide. It might therefore be asked why the unbiased interval should be chosen. The risk of the interval ρ_1 to ρ_2 failing to cover the true population value of ρ is fixed in both cases as 0·05, and since the unbiased interval is the greater over a very large range of r, why not choose the other? The answer is found in the definition of bias. In the case of the unbiased interval we see that this interval is chosen to cover the true population value 95 times in 100, and any other wrong value fewer times. In the case of the equal tail areas the interval is chosen to cover the true population value 95 times in 100, but it may cover some other wrong value more times than it does the true value. In the case of the r-distribution this discussion is, of course, theoretical, since the bias is so small, but it is conceivable that the point will prove important in other distributions.

It was expected that the bias in the distribution of r would prove to be small, because of Prof. Fisher's z' transformation (Fisher, 1921) for r. This transformation is nearly perfect, and transforms the asymmetrical curves of r into a series of normal curves. We should take equal tail areas from these normal curves in order to get unbiased limits for z', and therefore on transforming back we should expect to take equal tail areas from the r-curves. Since the transformation is not quite perfect we should expect a slight bias, which is what is actually found.

REFERENCES

NEYMAN, J. (1934). *J. R. statist. Soc.* XCVII, 589–93.

NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* I, 1–37.

DAVID, F. N. (1937) "Tables of the Ordinates and Probability Integral of the distribution of r in small samples." (Shortly to be published.)

FISHER, R. A. (1921). *Metron*, I, Part 4, 14.

BIOMETRIKA PUBLICATIONS: BOOKS OF TABLES

Tables for Statisticians and Biometricians

Edited by Karl Pearson

Part I

83 pages of Introduction and 143 pages of Tables

Price: 15s. net, packing and postage, Inland 9d., Abroad 1s. 6d.

Part II

250 pages of Introduction and 262 pages of Tables

Price: 33s. net, packing and postage, Inland 9d., Abroad 2s.

Tables of the Complete and Incomplete Β-Function

Edited by Karl Pearson

59 pages of Introduction and 494 pages of Tables

Price: 55s. net, packing and postage, Inland 1s., Abroad 2s.

Tables of the Incomplete Γ-Function

Edited by Karl Pearson

31 pages of Introduction and 164 pages of Tables

Price: 42s. net, packing and postage, Inland 9d., Abroad 1s. 3d.

Tables of the Complete and Incomplete Elliptic Integrals

(from Legendre's *Traité des Fonctions Elliptiques*. With autographed portrait of Legendre)

39 pages of Introduction by Karl Pearson and 94 pages of Tables

"Not the least valuable part of the book is the excellent introduction written by Prof. Karl Pearson. Here we have 39 pages devoted to a lucid explanation of the Tables and their use. Several interesting diagrams are shown and a beautiful autographed portrait of Legendre is included. As the Tables are a photographic reproduction the book is larger than usual, but all those whose work necessitates the use of the Tables will be grateful to the *Biometrika* Office for issuing such a valuable work of reference at so reasonable a price." NATURE

Price: 12s. 6d. net, packing and postage, Inland 9d., Abroad 1s. 3d.

Tests of Normality

By R. C. Geary and E. S. Pearson

This 16 pp. brochure reproduces the tables and charts which have already appeared in *Biometrika* together with some introductory matter containing illustrative examples.

Price 1s., post free

TO BE OBTAINED DIRECTLY FROM

The *Biometrika* Office, University College, London, W.C. 1

or through any Bookseller

At Press

NEW *BIOMETRIKA* PUBLICATION

Tables of the Ordinates and Probability Integrals of the Normal Correlation Coefficient

By F. N. DAVID

The inadequacy of the Standard or Probable Error of the Normal Surface Correlation Coefficient in cases where the sample is small, or the sample being large the coefficient is large, has long been recognised. These Tables are designed to assist the investigator, (1) in determining the probability that a sample or samples have been taken from a parent population of known correlation, and (2) in estimating, with a given measure of confidence, the limits within which this population correlation lies. Charts, auxiliary to the Tables, will make possible rapid extraction of significance levels and 'confidence' or 'fiducial' intervals.

PUBLICATIONS OF THE DEPARTMENT OF STATISTICS
UNIVERSITY OF LONDON
UNIVERSITY COLLEGE

Statistical Research Memoirs

Edited by J. NEYMAN and E. S. PEARSON

VOLUME I. June 1936

CONTENTS

1. Contributions to the theory of testing statistical hypotheses. I. Unbiased critical regions of Type A and Type A_1. By J. NEYMAN and E. S. PEARSON.
2. An investigation into the application of Neyman and Pearson's L_1-test, with Tables of percentage limits. By P. P. N. NAYER.
3. Note on an extension of the L_1-test. By B. L. WELCH.
4. Tests of certain linear hypotheses and their application to some educational problems. By PALMER O. JOHNSON and J. NEYMAN.
5. On the analysis of k samples from exponential populations with especial reference to the problem of random intervals. By P. V. SUKHATME.
6. Sufficient statistics and uniformly most powerful tests of statistical hypotheses. By J. NEYMAN and E. S. PEARSON.
7. Tests of statistical hypotheses in the case when the set of alternatives is discontinuous, illustrated on some genetical problems. By ROBERT W. B. JACKSON.

Subscription price, in advance, 12s. 6d. *per volume. Wrapper price* 15s.

OBTAINABLE FROM

The Department of Statistics, University College, London, W.C. 1

TRACTS FOR COMPUTERS, XXI

Logarithmetica Britannica

A Standard Table of Logarithms to Twenty Decimal Places

By A. J. THOMPSON, Ph.D.

Part III. Logarithms of Numbers 30,000–40,000

This part, the last but one to be issued, is now available, price 15s. *net*

Cambridge University Press, Fetter Lane, London, E.C. 4

At press or in preparation

TRACTS FOR COMPUTERS, V. Revised and enlarged edition

Table of the Coefficients of Everett's Central Difference Interpolation Formula

By A. J. THOMPSON, Ph.D.

These Tables, with an argument interval of 0·001 for θ, have been re-arranged so that the coefficients corresponding to θ and $\phi\,(=1-\theta)$ are on opposite pages. Coefficients, ϵ_8, for the eighth differences have been added to the original table, and two additional tables are given containing :

(i) the exact values of ϵ_2, ϵ_4, ϵ_6 and ϵ_8, and their 2nd, 4th, 6th and 8th differences for an argument interval of 0·01;

(ii) the exact values of ϵ_{10} and ϵ_{12} and their even differences for an argument interval of 0·1.

The Introduction, which Dr Thompson is re-writing, will give illustrations of problems in which the new Tables will be found useful.

TRACTS FOR COMPUTERS, XXII

Logarithmetica Britannica

By A. J. THOMPSON, Ph.D.

Part II. Logarithms of Numbers 20,000–30,000

TRACTS FOR COMPUTERS, XXIII

Tables of $\tan^{-1}x$ and $\log(1+x^2)$

To assist in the calculation of the ordinates of a Pearson Type IV curve

By L. J. COMRIE, Ph.D.

Statistical Research Memoirs

Edited by J. NEYMAN and E. S. PEARSON

VOLUME II. 1938

This volume will include some or all of the following papers:

A theorem concerning linearity of regression. By H. V. ALLEN.

Limiting distributions connected with problems of sampling human populations. By F. N. DAVID.

An extension of the Markoff theorem on least squares. By F. N. DAVID and J. NEYMAN.

On the best unbiased quadratic estimate of the variance. By P. L. HSU.

Problems of testing hypotheses and of estimation, concerning the sensitivity of psychological tests. By ROBERT W. B. JACKSON.

An investigation into the power function of the L_1 test for differences in standard deviations. By U. S. NAIR.

Contributions to the theory of testing statistical hypotheses. II. Theory of unbiased tests regarding more than one parameter. By J. NEYMAN and E. S. PEARSON.

On the power function of the z test. By P. C. TANG.

STATISTICAL STUDIES IN GENETICS AND HUMAN INHERITANCE

ANNALS OF EUGENICS

EDITED BY R. A. FISHER

Vol. VIII, Part I

CONTENTS

William Fleetwood Sheppard, 1863–1936. Frontispiece.

Obituary. N. F. SHEPPARD.

A note on Sheppard's contribution to mathematics and mathematical statistics. A. C. AITKEN.

The character of Sheppard's work. R. A. FISHER.

Bibliography.

Studies on a child population. II. Retests on the advanced Otis and Stanford-Binet scales, with notes on the use of a shortened Binet scale. J. A. FRASER ROBERTS and R. GRIFFITHS.

Sexual and racial variations in ability to taste Phenyl-Thio-Carbamide, with some data on the inheritance. W. C. BOYD and L. G. BOYD.

Tests of significance for samples of the χ^2-population with two degrees of freedom. P. V. SUKHATME.

Significance of grouping. W. L. STEVENS.

A test for uniovular twins in mice. W. L. STEVENS.

Inheritance in man: Boas's data studied by the method of analysis of variance. R. A. FISHER and H. GRAY.

The test of significance for Mendelian ratios when classification is uncertain. H. F. SMITH.

The analysis of single-factor segregations. K. MATHER.

Vol. VIII, Part II

CONTENTS

A note on the importance of differences between examiners in the estimation of the ability of students. J. B. HUTCHINSON and B. M. PUGH.

The stature of male schoolchildren in relation to hair and eye colour. M. A. MACCONAILL.

The sex incidence of certain hereditary traits in man. L. CSIK and K. MATHER.

Maximum likelihood, inverse probability, and the method of moments. H. JEFFREYS.

Blood groups and genetics. V. FRIEDENREICH.

Familial myoclonus epilepsy associated with deaf-mutism in a family showing other psychobiological abnormalities. A. D. LATHAM and T. A. MUNRO.

Note on the distribution of the three forms of *Lythrum Salicaria*. D. MCGREGOR. Editorial note.

Studies on a child population. III. Intelligence and family size. J. A. FRASER ROBERTS, R. M. NORMAN and R. GRIFFITHS.

Reviews.

Subscription in advance, 50s. per volume. Four quarterly parts obtainable separately at 15s. each, from

THE GALTON LABORATORY, UNIVERSITY COLLEGE, LONDON, W.C.1

THE ANNALS OF MATHEMATICAL STATISTICS

Vol. VIII, Number 2. June 1937

CONTENTS

Regression and Correlation Evaluated by a Method of Partial Sums. FELIX BERSTEIN.

Methods of Obtaining Probability Distributions. BURTON H. CAMP.

Moment Recurrence Relations for Binomial, Poisson and Hypergeometric Frequency Distributions. JOHN RIORDAN.

Note on Zoch's Paper on the Postulate of the Arithmetic Mean. ALBERT WERTHEIMER.

Note on the Binomial Distribution. C. E. CLARK.

Convexity Properties of Generalized Mean Value Functions. NILAN NORRIS.

A Simple Form of Periodogram. DINSMORE ALTER.

Vol. VIII, Number 3. September 1937

CONTENTS

On Certain Distributions Derived from the Multinomial Distribution. SOLOMON KULLBACK.

A Problem in Least Squares. JAN K. WISNIEWSKI.

A Significance Test for Component Analysis. PAUL G. HOEL.

Contributions to the Theory of Comparative Statistical Analysis. I. Fundamental Theorems of Comparative Analysis. WILLIAM G. MADOW.

Reply to Mr Wertheimer's Paper. RICHMOND T. ZOCH.

Published Quarterly. Four Dollars per annum

Post Office Box 171, Ann Arbor, Michigan, U.S.A.